

Variation within Optimality Theory

**Proceedings of the Stockholm Workshop on
'Variation within Optimality Theory'**

**April 26-27, 2003
at Department of Linguistics
Stockholm University
Sweden**

**Eds:
J. Spenader, A. Eriksson & Östen Dahl**

Foreword

In pace with the widening range of problems that Optimality Theory has been applied to, the different versions or the theory, or 'dialects' as we have called them, have been growing as well. However, it is not quite clear what the virtues and vices of each of the OT-dialects are when applied to different problems, and the desire to have some focused discussion on these issues was what inspired the Stockholm workshop.

We were very pleased by the response for the workshop, receiving 20 proposals for presentations. Each proposal was reviewed by at least two members of the programme committee (Paul Boersma, Anders Eriksson, Östen Dahl, Hanjung Lee, Tomas Riad, Jennifer Spenader and Henk Zeevat). Special thanks also to Torgrim Solstad for some additional reviewing. In addition to submitted abstracts, a large number of researchers registered their intention to participate weeks before the workshop, showing that there seems to be a great interest in the topics that will be covered. The table of contents shows clearly that researchers from all areas of linguistics: phonology, morphology, syntax, semantics, pragmatics and computational linguistics, have something to say about how different versions of OT measure up.

Finally we would also like to gratefully acknowledge the support of Kungl. Vitterhets Historie and Antikvitets Akademien in financing the visit of Dr. Paul Boersma, our invited speaker, and the Department of Linguistics at Stockholm University for supporting the workshop both financially and practically.

14 April, 2003
Stockholm University, Sweden

Jennifer Spenader, Anders Eriksson & Östen Dahl

Original Call for Papers

Recently there has been a proliferation of different "dialects" of optimality theory (OT); e.g. bi-directional optimality theory, stochastic optimality theory, primitive optimality theory, etc. Many of these dialects were developed to handle short-comings in standard OT for problems particular within a specific linguistic field, but it is not clear how the different OT dialects work for problems outside that particular area. This workshop aims to bring together researchers using different forms of OT in different fields within linguistics, including phonetics, phonology, morphology, syntax, semantics and pragmatics. The emphasis is on how different OT dialects support or fail to support the analysis of certain problems in order to make their differences and similarities more transparent. The characteristics of the different forms of OT and how they relate to different problems, rather than the characteristics of the analysed problems themselves, should be the central focus. We invite abstracts on all topics related to optimality theory, including, but not limited to:

- comparisons between different forms of OT
- comparative studies of the same problem within more than one form of OT
- application of an OT-dialect to a problem in a field new to that dialect
- discussions of the inability of some forms of OT to handle certain problems
- discussions of the meta-characteristics of the different types of OT
- discussions of learning algorithms for different types of OT and how they measure up with different data
- discussions of computer implementations of OT dialects and their characteristics

In addition to talks we may also make time for demonstrations of computer implementations of OT-algorithms.

OT variations in phonology and morphology

Towards an optimal account of diachronic chain shifts: Part I (Grimm's law)

Sang-Cheol Ahn,
Kyung Hee University

This paper proposes a new account on Grimm's law within the dispersion version of Optimality Theory. I first argue that the notion of markedness should be employed to trace the trigger of the whole change, while ease of articulation can also be a crucial factor accounting for subsequent changes. Then, I show that an Optimality-theoretic approach employing Dispersion Theory (Flemming 1995, 1996) provides a natural account on those historical changes. Here, I claim that the overall shift can be explained better in terms of pattern evaluation since all the changes are related to each other, obeying the "no merge" principle. Furthermore, I claim that the differences in the changes according to the historical stages can be accounted for with respect to constraint conjunction, rather than different constraint ranking.

1. Introduction

1.1. Grimm's law

In the Proto-Indo-European consonant system, there were three types of stops, voiced aspirates /b^h, d^h, g^h, g^{hw}/, voiced stops /b, d, g, g^w/ and voiceless stops /p, t, k, k^w/, while only one fricative /s/ was existent in the phonemic inventory. During the development of Germanic languages being separated from other Indo-European, however, the stops underwent massive chain shifts, following the so-called Grimm's law. Here, setting aside the fricative /s/, we observe two major facts. First, by Grimm's law, three types of stops formed a chain shift; [+voice, + aspirated] > [+voice, -aspirated], [+voice, -aspirated] > [-voice, -aspirated], [-voice, -aspirated] > [-voice, +continuant], as shown in the following table. Second, the labio-velar stops lost the labial articulation feature and thus merged with the velar stops.

(1)	<i>Proto-Indo-European</i> ¹		<i>Germanic</i>
	b ^h , d ^h , g ^h , g ^{wh}	>	b, d, g
	p, t, k, k ^{wh}		f, θ, x(h)
	b, d, g, g ^w		p, t, k

¹ Recently, however, a number of linguists adopted a different proposal on the PIE phonemic system, known as the "Glottalic Theory". The proposal is that the PIE series traditionally reconstructed as voiced plosives /b, d, g, g^w/ was actually an ejective series /p', t', k', k^{w'}/, which would explain why the segment /p'/ was rare or absent. Moreover, the former /p, t, k, k^w/ are interpreted as /p^h, t^h, k^h, k^{hw}/ (Trask 1996: 233-235).

Accounting for these historical changes, numerous studies have proposed the following types of rules.

- (2) a. [+voice, -continuant] → [-voice]
 b. [+voice, +aspirated] → [-aspirated]
 c. [-voice, -aspirated] → [+continuant]

We, however, can note that, within this sort of a rule-based account, it is difficult to provide any explanatory description on these changes. In other words, being context-free, these rules cannot explain the initial cause or the sequence of the changes.

On the other hand, Pyles & Algeo (1993: 90) argue that each set of the changes was completed before the next began. In the accompanying table, therefore, Pyles & Algeo (1993) number the steps in the order in which they happened. (The missing number [3] is the change described as Verner's law shifting voiceless fricatives to voiced ones.)

- (3) $\begin{array}{llll} b^h, d^h, g^h & [1] & > & \beta, \delta, \gamma & [5] & > & b, d, g \\ p, t, k & [2] & > & f, \theta, x & (h \text{ initially}) \\ b, d, g & [4] & > & p, t, k \end{array}$

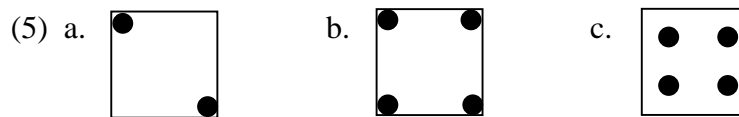
In this description, however, it is difficult to find any theoretical (or textual) evidence for the steps of the changes. For example, there was no explanation on the cause of the initial change, $/b^h, d^h, g^h/ > / \beta, \delta, \gamma/$. Similarly, it is difficult to find a cause changing the intermediate voiced fricatives to voiced stops. In other words, there is no reason for the dental fricative $/ \delta/$ to become a stop $/d/$. Moreover, observing that syllable-initial aspiration is very common in modern Germanic languages, it is quite questionable to assume that the plain voiceless stops underwent spirantization, without undergoing an intermediate stage, i.e., the aspirated stops $/p^h, t^h, k^h/$. Finally, earlier studies did not explain why the labialized stops disappeared.

1.2. *Dispersion Theory*

In Optimality Theory (OT henceforth, McCathy & Prince 1995), we allow all possible candidate outputs and then evaluate them with a set of relevant constraints. The main analytical proposal of OT is that constraints are ranked in a hierarchy of relevance. Lower-ranked constraints can be violated in an optimal output to respect higher-ranked constraints. An optimal output can thus minimally violate certain low-ranked constraints. In Dispersion Theory, on the other hand, there are constraints on the well-formedness of phonological contrasts. Specifically, the selection of phonological contrasts is subject to the following three functional goals (Flemming 1995, 1996).

- (4) a. Maximize the number of contrasts.
 b. Maximize the distinctiveness of contrasts.
 c. Minimize articulatory effort.

The possibility of incorporating these principles into OT emerges from the fact that the functional goals in (4) are in conflict with each other. The following figures illustrate the relations among the three requirements (Flemming 1995, 1996).



(5a) shows an inventory including only one contrast, but the contrast is maximally distinct since the two sounds are far apart from each other in the auditory space. (5b) shows the case in which we fit more sounds into the same auditory space since the sounds are closer together here. Therefore, the goals of maximizing the number of contrast and maximizing the distinctiveness of contrast conflict. Moreover, the third constraint for ease of articulation also conflicts with the constraint maximizing distinctiveness. As the sounds in the periphery of the space requires more effort than those located in the less peripheral regions, it is necessary to restrict sounds to a reduced area as shown in (5c).

The basic notions of Dispersion Theory can be incorporated in the framework of OT in that the requirements on contrast conflict and the selection of an inventory of contrast involves achieving a balance between them (Flemming 1996, 2001). In this paper, therefore, it will be claimed that the final output of the obstruent system is a consequence of interactions among several phonetically natural constraints. As the well-formedness of the consonantal system cannot be evaluated in isolation, the overall result is obtained by the pattern evaluation of the adjacent consonants.

2. On trigger and subsequent changes

2.1. *Trigger*

We assume the following four types of PIE obstruents (Iverson & Salmons 2001).

(6)

Stops			Fricative
b^h, d^h, g^h, g^{hw}	b, d, g, g^w	p, t, k, k^w	s
[+voice], [+aspirated]	[+voice]		[+continuant]

Here we observe that the PIE stop system was highly marked in that it required both voicing and aspiration, which made the “**weak**” point initiating a chain shift. Therefore, we can invoke an inviolable constraint suppressing voiced aspirated stops, which could have triggered the whole shift: i.e., those voiced aspirated stops had to change to other types of consonants. From a purely conjectural point of view, however, we may consider a couple of possible paths for the change since they could have undergone spirantization. First, we may think about the notion of “**ease of articulation**” for spirantization, assuming that the spirants (i.e., fricatives) require less effort than stops. This option, however, is highly speculative (Anttila 1972: 189). If spirants are easier to pronounce than stops, it is difficult to explain why the fricatives became stops at the final stage of Grimm’s law (as argued in Pyles & Algeo (1993)). As Anttila (1972) admits, it is very unusual to have voiced spirants without voiceless ones in a language having the feature [voiceless]. Also, as shown in (7), why would the Baltic Finnic speakers have replaced them with stops?

(7) Consonant correspondences between Germanic (English) loans in Baltic Finnish

<i>English (Germanic)</i>	<i>Finnish (Baltic Finnish)</i>
/f/ field, Friday	/p/ pelto, perjantai

/θ/	death, (Gothic) aipei	/t/	tauti 'sickness', äiti 'mother'
/h/	hen	/k/	kana

These arguments indicate that it is more probable for the voiced aspirated stops to become the plain voiced stops (Iverson & Salmons 2001). Thus, within OT, we need a constraint like Ident[cont] to discourage the voiced aspirated stops not to undergo spirantization. Then all the stops should remain as stops. Moreover, observing that those labio-velar stops /g^w, g^{wh}, k^w/ disappeared in an earlier stage, we can posit another inviolable constraint *Complex(Place) prohibiting a segment with complex articulation.

- (8) Ident[cont]: The [continuant] feature of the input may appear in the output.
 *Complex(Place): Segments with complex place of articulation are suppressed.

Employing these constraints, the following step is proposed as the initial change of Grimm's law.

- (9) Step 1: b^h, d^h, g^h, g^{hw} > b, d, g, (g)

2.2. The subsequent changes

The first stage of the change now forces the original voiced stops to become the voiceless stops because the voiced stops merged with the original voiceless stops. Here we need the notion of "**pattern evaluation**" since the changes of the single segments cannot be considered separately; they are evidently parts of one great linguistic movement. Following Flemming (1996), therefore, I propose the following constraint avoiding possible merger.

- (10) Maintain Contrast
 The phonemic contrast of the input should be maintained in the output.

Then the second stage of the change should have been the /b, d, g/ > /p, t, k/ pattern.

- (11) Step 2: b, d, g > p, t, k

The steps 1 and 2 show that the change of the PIE obstruents was triggered and succeeded by the reduction of the marked values, i.e., [+voice, +aspirated] > [+voice], [+voice] > [-voice].

The next target of the chain shift is the voiceless stops /p, t, k/. Note, however, that they do not follow the general scheme of the earlier two stages, i.e., reduction of markedness since they had to become fricatives or aspirated stops. The voiceless stops could not change to voiced stops since the original slots for the voiced stops have been taken by the voiced aspirates. Then, there are a couple of options for their changes. First, they could have change to fricatives, i.e., /p, t, k/ > /f, θ, x (h)/, as often claimed in earlier literature (King 1969, Anttila 1972, Trask 1996, Pyles & Algeo 1993, etc.). This possibility, however, is quite unnatural from a phonetic point of view in that the phonetic quality of the plain stops is quite distinct from that of the fricatives. If allowed, it has to be an abrupt change, while most historical changes tend to be quite **gradual**. Moreover, in terms of pattern symmetry, there is a good reason for the voiceless fricatives not to undergo spirantization. Note that the Proto-Indo-European system had only one fricative /s/ in the phonemic inventory. Thus, if the voiceless stops underwent spirantization di-

rectly, the following asymmetric phonemic system with two similar (i.e., coronal) fricatives /θ, s/ might have appeared.

- (12)
- | | | | |
|---|---|---|---|
| p | | t | k |
| b | | d | g |
| f | θ | s | x |

So, the question arises: why did /t/ become the interdental fricative /θ/, changing the alveolar value of the stop? Or, why didn't /t/ merge with /s/ during the change?

Iverson & Salmons (2001) argue that in the speech community destined to become Germanic, phonological developments began with the introduction of aspiration into the ancestral voiceless stops. They term this key innovation to the IE obstruent system "Germanic enhancement" and see it as a catalyst for extensive subsequent change.

- (13) Germanic enhancement: Laryngeally unspecified stop → [+spread glottis]

As they admit, however, it is quite speculative to claim that Germanic enhancement was the trigger of the whole shift. Nevertheless, it is a plausible argument that the voiceless stops became aspirated before undergoing spirantization in the end. Thus, taking this generalization, we can argue that the voiceless plain stops became aspirates due to Germanic enhancement.

- (14) Step 3: /p, t, k/ > /p^h, t^h, k^h/

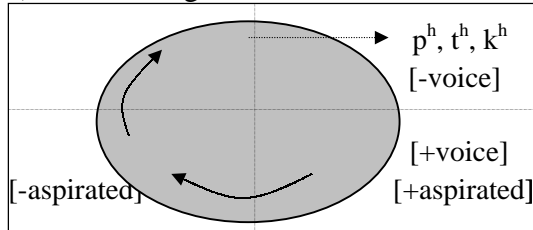
We can provide at least two arguments for this assumption. First, we might note that this generalization, Germanic enhancement still persists in the phonetic system of most modern Germanic languages. For example, English and German show the distinction between aspiration in syllable initial stops vs. no aspiration in other environments, while a similar aspect is realized as preaspiration in Icelandic. But it is difficult to find such aspiration in other languages like Romance or Slavic languages.

- (15) { Aspiration: English, German, Swedish, Norwegian, Danish, etc.
 { Preaspiration: Icelandic (cf. No aspiration: Dutch, Yiddish)

The second evidence comes from the orthographic representation in Modern English. As [f] is often transcribed as *ph* in English, while *f* in Romance languages, which seems to indicate that PIE /p/ went first to Proto-Germanic /f/, parallel to the non-strident outputs of PIE /t, k/-spirantization. This orthographic evidence is indirect but can be used as a clue indicating that aspiration of the voiceless stops could have occurred.

Adopting the view of the change /p, t, k/ > /p^h, t^h, k^h/, this process is different from the earlier steps of change, i.e., reduction of markedness, since aspirated stops are more marked than plain stops. Rather, it is a "**weakening**" process as aspirated stops are weaker than equivalent unaspirated ones, and their briefer closure durations are more susceptible to becoming incomplete (Hooper 1976: 224). Based on the arguments made so far, therefore, we can provide the following figure showing that the initial stage was triggered by the deaspiration of the voiced aspirates, causing a "**push-chain**" type of successive changes. The shaded area represents the earliest PIE obstruent system triggered by the principle, reduction of markedness, while the dotted arrow shows the last step, i.e., the beginning of the weakening process.

(16) Initial stages of Grimm's law

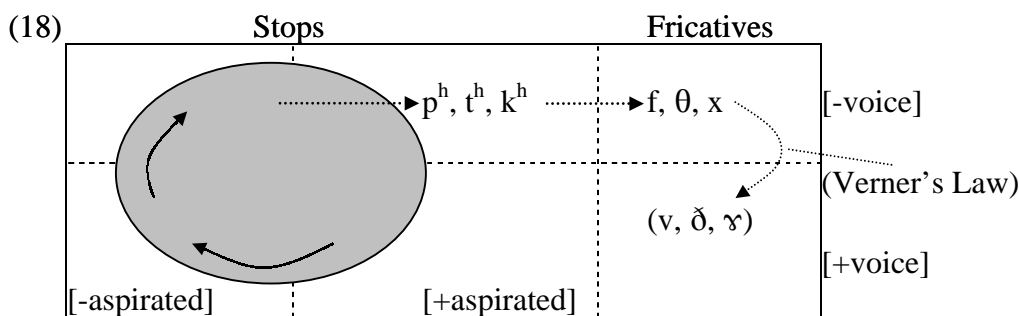


As the last step in (16) shows the weakening process, this process went on further, producing spirantization. In other words, once the original voiceless stops became aspirated via Germanic Enhancement, the period of the stop occupied by the closure became less as the period occupied by the voiceless release (aspiration) became greater, assuming a more or less constant overall duration of the stops. Then, as spirants are weaker than aspirated stops, those aspirated stops undergo spirantization in a later stage.

(17) Step 4: /p^h, t^h, k^h/ > /f, θ, x/

Iverson & Salmons (2001) regards this as a case of “hyper-enhancement”, namely, the spirantization of phonetically aspirated stops. This change is akin to the later changes associated with the High German Shift or the incipient affrication of aspirated stops currently underwent in Danish. With spirantization of aspiration enhanced voiceless stops in late Indo-European/early Germanic, the original fricative class expanded considerably (and compensatorily, à la Iverson & Salmons (2001)) as the contrastive stops reduced from the three of Indo-European to the two of Germanic.

Therefore, I assume that the later stages of Grimm's law, i.e., steps 3 and 4, were the consequences of the weakening process for ease of articulation. We can sum up the whole process of Grimm's law as follows. (The original PIE stops (except the labio-velar stops, for convenience) are enclosed in the shaded area, while the later stages are shown in the white area.)



The solid arrows are used for the “marked → unmarked” changes, while the dotted arrows for weakening (due to ease of articulation). Moreover, as the parentheses indicate, the later stage of weakening process went further as Verner's law. As a consequence, we can also categorize the various stages of Grimm's law as follows.

(19) Principles in Grimm's law

- a. Reduction of markedness: /b^h, d^h, g^h/ > /b, d, g/, /b, d, g/ > /p, t, k/
- b. Ease of articulation
 - i. Weakening: /p, t, k/ > /p^h, t^h, k^h/, /p^h, t^h, k^h/ > /f, θ, x/
 - ii. Simplification of complex articulation: /g^w, g^{hw}, k^w/ > /g, g, k/

3. OT account

3.1. Pattern evaluation

We have observed that the initial cause of the change was the reduction of markedness which changed the aspirated stops /b^h, d^h, g^h/ to the plain /b, d, g/. Therefore, within OT, the initial change seems to be easily accounted for by the following constraint.

(20) *Asp(iration): Aspirated segments may not be allowed.

Being a more general shape, *Asp should take a major role forcing the intermediate voiceless aspirated stops to undergo spirantization. As shown in the following tableau, the trigger constraint *Asp takes the crucial role forcing the deaspiration of /b^h, d^h, g^h/, while Ident[voice] eliminates the competing candidates, voiceless aspirates /p^h, t^h, k^h/. Moreover, two faithfulness constraint Ident[cont] and Ident[voice] are also inviolable.

(21) Step 1 (Initiation of Grimm's law): /b^h/ > /b/

/b ^h /	*Asp	Ident[cont]	Ident[voice]	Ident[asp]
a. b ^h	*!			
b. p ^h	*!		*	
c. v		*!		*
d. p			*!	*
e. b				*

In the next step of the change, however, we need further consideration since we might get the same type of stops /b, d, g/ which remain unchanged. As mentioned in the earlier section, the whole chain shift was caused by the reduction of markedness and the subsequent changes were made to avoid possible merges. Thus, we need to employ the mechanism of “**pattern evaluation**” of Dispersion Theory, in which all the possible input-output correspondence candidates should be evaluated in conjunction with other groups of candidates since all the changes of chain shift are evaluated are tied up with each other. Adopting this mechanism, we employ Maintain Contrast constraint invoked earlier, which is now ranked the highest to prevent possible merges between new outputs and those from the earlier process. Note that, being violated by the optimal candidate /p/, Ident[voice] now becomes violable. (Here the parentheses show the output segments (and their violation marks) from the earlier, i.e., the initial change.)

(22) Step 2: /b/ > /p/ (pushed by /b^h/ > /b/)

/(b ^h) b/	Maintain Contrast	*Asp	Ident[cont]	Ident[voice]	Ident[asp]
a. (b ^h) b ^h	*!	*!(*)			*
b. (b) b	*!				(*)
c. (p) b ^h		*!		(*)	*
d. (p ^h) v		(*)	*!		
e. (b) p				*	(*)

3.2. Local conjunction

As we move to the next stage, however, we find a new problem in that the /p/ > /p^h/ change should violate the high ranking *Asp constraint. Therefore, we may consider a

new constraint ranking demoting *Asp, in order to allow “weakening” (i.e., minimal violation of *Asp). Although we demoted *Asp to the bottom, however, we get the incorrect candidate (23c) as the optimal output, rather than the correct (23d). (Here the parentheses also show the outputs and their violation marks of the earlier changes.)

(23) Step 3: /p/ > /p^h/ (pushed by /b/ > /p/)

	/b ^h b/	p/	Maintain Contrast	Ident[cont]	Ident[voice]	Ident[asp]	*Asp
a.	(b ^h p)	b ^h	*!		*(*)	*	*(*)
b.	(b p)	p	*!		(*)	(*)	
☛ c.	(p ^h b)	b ^h			*(*)	*	*(*)
? d.	(b p)	p ^h			(*)	*(*)	*
e.	(b p)	b ^h			*(*)	*(*)	*(*)

In order to trace the fundamental problem, therefore, we go back to the initial analysis for the triggering stage. The reason for proposing the general form of a constraint, *Asp, was to make it trigger the “marked > unmarked” /b^h/ > /b/ change in the initial stage, while allowing aspiration (i.e., weakening) in the later stage /p/ > /p^h/ by the demotion of *Asp. We should thus take a closer look at the triggering factor of the chain shift, observing that the motivation of the change was to eliminate the voiced aspirates. And those earlier voiced aspirates have never showed up in Germanic languages since the initial change. Therefore, we need to conjoin two constraints, *Voice and *Asp to prevent the voiced aspirates in Germanic. Note that the first two steps of the changes were motivated by the reduction of markedness, *Voice and *Asp are the well-motivated constraints discouraging marked values. Moreover, as will be shown below, *Voice takes an important role preventing voiced fricatives in the later stage, /p^h, t^h, k^h/ > /f, θ, x/. The local conjunction of constraints now takes a crucial role in both initial and later stages of Grimm’s law.

(24) Step 1 (Initiation of Grimm’s law): /b^h/ > /b/

	/b ^h /	*Asp&*Voice	Ident[cont]	Ident[voice]	Ident[asp]	*Voice	*Asp
a.	b ^h	*!				*	*
b.	p ^h			*			*
c.	v		*!		*	*	
d.	p			*!	*		
☛ e.	b				*	*	

(25) Step 3: /p/ > /p^h/ (pushed by /b/ > /p/)

	/b ^h b/	p/	*Asp&*Voice	Maintain Contrast	Ident [cont]	Ident [voice]	Ident [asp]	*Voice	*Asp
a.	(b ^h p)	b ^h	*!(*)	*!		*(*)	*	*(*)	*(*)
b.	(b p)	p		*!		(*)	(*)	(*)	
c.	(p ^h b)	b ^h	*!			*(*)	*	*(*)	*(*)
☛ d.	(b p)	p ^h				(*)	*(*)	(*)	*
e.	(b p)	b ^h	*!			*(*)	*(*)	(*)	*(*)

Due to the inviolable role of the conjoined constraint, we do not need constraint re-ranking which has been criticized by McMahon (2000). Moreover, the local conjunction avoids the possible logical problem argued in Kager (1999) since the role of the conjoined constraint is to eliminate the worst of the bad candidates. *Asp and *Voice

play independent roles in various stages of Grimm's law but they are ranked relatively low since their violation is not fatal. Furthermore, we can explain the cause of the so-called Germanic enhancement in a more natural way.

When we move to the last step of Grimm's law, $/p^h, t^h, k^h/ > /f, \theta, x/$, we face a new problem in that the noncontinuancy of the stop is to be violated in this step. As this process is pushed by the previous step $/p, t, k/ > /p^h, t^h, k^h/$, we need to take both process into consideration in pattern evaluation.

(26) Step 4: $/p^h/ > /f/$ (pushed by $/p/ > /p^h/$)

$/(b^h \ b) \ p^h/$	*Asp&*Voice	Maintain Contrast	Ident [cont]	Ident [voice]	Ident [asp]	*Voice	*Asp
a. $(b^h \ p) \ p^h$	*!			(*)		(*)	*(*)
? b. $(b \ p) \ f$			*	(*)	(*)	(*)	
c. $(b \ p) \ v$			*	*(*)	*	*(*)	
☛ d. $(p^h \ b) \ p$				(*)	*	(*)	(*)
e. $(p \ b) \ b^h$	*!			*(*)	*	*(*)	*

Although we can eliminate the wrong candidate $/v/$ in (26c) from the competition with the correct output $/f/$ (26b), we still have to eliminate the expected incorrect output $/p/$ in (26d). We might consider another local conjunction, such as *Asp&*Voiceless to escape from this difficulty. This conjunction, however, causes a more serious problem in that it should prevent the voiceless aspirates in the earlier stage as we do not adopt constraint re-ranking. Moreover, more local conjunction would make the whole grammar more complicated. Considering that this is another case of weakening, therefore, we need the following constraint enforcing voiceless stops to spirantize.

(27) *Strengthen[-voice]: Voiceless segments may not undergo strengthening in the process of a historical shift.

This constraint does not allow the voiceless aspirated stops to go back to the previous stage by undergoing strengthening.

(28) Step 4: $/p^h/ > /f/$ (pushed by $/p/ > /p^h/$)

$/(b^h \ b) \ p^h/$	*Asp&*Voice	*Strengthen [-voice]	Maintain Contrast	Ident [cont]	Ident [voice]	Ident [asp]	*Voice	*Asp
a. $(b^h \ p) \ p^h$	*!				(*)		(*)	*(*)
☞ b. $(b \ p) \ f$				*	(*)	(*)	(*)	
c. $(b \ p) \ v$				*	*(*)	*	*(*)	
d. $(p^h \ b) \ p$		*!			(*)	*	(*)	(*)
e. $(p \ b) \ b^h$	*!				*(*)	*	*(*)	*

Due to the role of the new constraint, therefore, we can get the select the correct candidate. Based on the discussion made so far, we can list the constraints and their ranking, regardless of the stages of the change.

(29) *Asp&*Voice, *Strengthen[-voice] >> Maintain Contrast >> Ident[cont] >> Ident[voice] >> Ident[asp] >> *Voice >> *Asp

References

- Ahn, Sang-Cheol. 2001. An optimality approach to chain shifts: Nasal vowel lowering in French. *Language Research* 37.2, 359-375.
- Ahn, Sang-Cheol. 2002a. A dispersion account on Middle Korean Vowel shifts. *Japanese/Korean Linguistics* 10, 237-250.
- Ahn, Sang-Cheol. 2002b. An optimality approach to the Great Vowel Shift. *Korean Journal of Linguistics* 27.2, 153-170.
- Anttila, Raimo. 1972. *An Introduction to Historical and Comparative Linguistics*. New York: McMillan.
- Avery, Peter and William Idsardi. 2000. Laryngeal dimensions, completion and enhancement. In *Distinctive Features* (Proceedings of the 1999 Zentrum für Allgemeine Sprachwissenschaft (ZAS) Conference on Distinctive Features, Berlin).
- Calabrese, Andrea and Morris Halle. 1998. Grimm's and Verner's Laws: a new perspective. In Jay Jasanoff, Craig Melchert & Lisi Olivier (eds.) *Mír Curad: Studies in Honor of Calvert Watkins*. Innsbruck: Institute für Sprachwissenschaft, University of Innsbruck, pp. 47-62.
- Flemming, Edward. 1995. *Auditory Representations in Phonology*. Doctoral dissertation, UCLA.
- Flemming, Edward. 1996. Evidence for constraints on contrast: the dispersion theory of contrast. *UCLA Working Papers in Phonology* 1, 86-106.
- Flemming, Edward. 2001. Contrast and perceptual distinctiveness. To appear in B. Hayes, R. Kirchner, and D. Steriade (eds.) *The Phonetic Bases of Markedness*. Cambridge University Press.
- Grimm, Jacob. 1826. *Deutsche Grammatik*. 2nd edition. Berlin: Bertelsmann.
- Hooper, Joan. 1976. *Natural Generative Phonology*. New York: Academic Press.
- Iverson, Gregory and Joseph Salmons. 1995. Aspiration and laryngeal representation in Germanic. *Phonology* 12, 369-396.
- Iverson, Gregory and Joseph Salmons. 1999. Glottal spreading bias in Germanic. *Linguistische Berichte* 178, 135-151.
- Iverson, Gregory and Joseph Salmons. 2001. Laryngeal enhancement in early Germanic. Paper presented at the 7th Germanic Linguistics Annual Conference, Banff, Alberta.
- Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.
- King, Robert D. 1969. *Historical Linguistics and Generative Grammar*. Englewood Cliffs, NJ: Prentice-Hall.
- Lass, Roger. 1997. *Historical Linguistics and Language Change*. Cambridge: Cambridge University Press.
- Lehmann, Winfred P. 1952. *Proto-Indo-European Phonology*. Austin: University of Texas Press.
- McCarthy, John and Alan Prince. 1995. Faithfulness and reduplicative identity. *University of Massachusetts Occasional Papers* 18: *Papers in Optimality Theory*, 249-384.
- McMahon, April. 2000. *Change, Chance, and Optimality*. Oxford University Press.
- Petrova, Olga. 2000. Grimm's law in Optimality Theory. *Proceedings of the Eleventh Annual UCLA Indo-European Conference*, 45-67.
- Trask, R. L. 1996. *Historical Linguistics*. London: Arnold.
- Verner, Karl. 1875. Eine Ausnahme der ersten Lautverschiebung. *Zeitschrift für vergleichende Sprachforschung* 23, 97-130.
- Weinstock, John. 1968. Grimm's Law in distinctive features. *Language* 44, 224-229.

The Deformity of Anti-Faithfulness

Diana Apoussidou

Institute of Phonetic Sciences, Amsterdam

Abstract. In Optimality Theory, an important tool for accounting for morpho-phonological processes is output-output correspondence. A development of this approach is Transderivational Anti-Faithfulness (TAF), constituting a reversal of faithfulness. This article will explore the nature of TAF, in order to test this approach as an extension of Optimality Theory. The example of morphologically triggered accent in Modern Greek will turn out to reveal some formal problems of TAF.¹

1. Introduction

If one has a look across paradigms, two observations can be made. One: most of the forms show a very similar structure. Two: many of the forms show a striking difference. This points toward different principles which stand in competition with each other. The similarity of forms in a paradigm can be seen as an ambition to express that these forms stand in relation to each other. The differences between them, however, code various information in these forms (for instance, suffixes can code the information of gender, person, or number, etc.). For the sake of intelligibility, this distinctive information should be expressed quite clearly. In Optimality Theory (OT, Prince & Smolensky 1993, McCarthy and Prince 1993), these two ambitions find their formalization in Transderivational Correspondence Theory (TCT, Benua 1997) and Transderivational Anti-Faithfulness (TAF, Alderete 1999, 2000). While TCT focuses on the similarity between forms of a paradigm, TAF focuses on the difference. As I want to show by the example of Modern Greek accent, Anti-Faithfulness encounters some formal problems.

The outline of this paper is as follows: in the next section, I illustrate the theory of output-output relations, TCT and TAF. In section 3, the phenomena which can be captured with Anti-Faithfulness will be described. Next, I will apply TAF to Modern Greek, a language that shows similar accent properties to the languages defined as involving TAF. It will turn out that Modern Greek poses some problems for the predictions made by TAF. In the concluding remarks, I will very shortly go into other fields where TAF was applied to, in order to discuss if TAF might contribute more to other phenomena than accent-related fields.

¹ The simplified notion of accent here involves stress as well as tone phenomena.

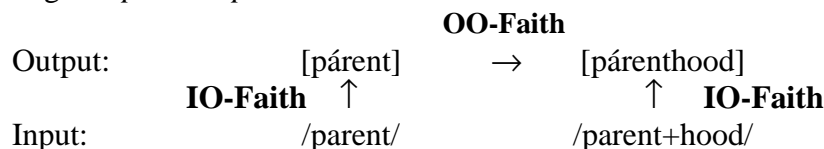
2. Output-Output Relations

While TCT (Benua 1997) accounts for similarities between morphologically related forms, TAF on the other hand is meant to account for the differences between morphologically related forms.

2.1 Transderivational Correspondence

TCT is an extension of OT that puts two output forms into correspondence with each other. With this approach, it is possible to explain the similarity between the forms of a paradigm. This is achieved through Output-Output-Faithfulness. The two corresponding strings are a base and a derivative form. Consider stress placement in English, where stress can stay in the same position, as shown in (1). The output forms *párent* and *párenthood* stand in connection with their input morphemes /parent/ and /-hood/, but they also stand in correlation with each other. OO-Faithfulness requires that these two outputs should be phonologically identical.² So in (1), stress is in both forms realized on the same syllable in both forms.

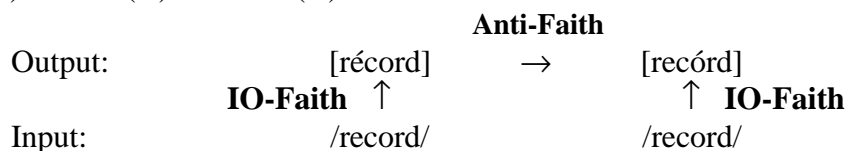
(1) English: *párent* - *párenthood*



2.2 Transderivational Anti-Faithfulness

Conversely, English morphologically related forms can also differ in their stress positions. Consider for instance verbs and nouns that are much the same regarding their segments, but that differ in their stress placement. A word like *record* is stressed on the initial syllable if it is a noun (*récord*), and stressed on the final syllable if it is a verb (*recórd*). Morpho-accentual processes like these can serve to strengthen the opposition between two morphological classes (2).

(2) *récord* (N) – *recórd* (V)



This opposition is in TAF expressed by means of Anti-Faithfulness constraints. This constraint class enforces the contrast between two morphologically related outputs, namely between a base and a related morphological derivative form. Anti-Faithfulness

² In this example, OO-Faithfulness refers only to accent placement. So the fact that the derivative form has more structure than the base, namely the segments /h/, /o:/, and /d/, and thus would violate a constraint like OO-DEP-STRUCTURE, is ignored here for the sake of simplicity.

constitutes a reversal of traditional Faithfulness, but is restricted to output-output relations. While Faithfulness as such seeks to maintain identity between two corresponding strings (3+4), Anti-Faithfulness is defined as requiring a difference in some respect (5).

- (3) General Faithfulness (McCarthy and Prince 1995)
 - MAXIMALITY: Every element of S_1 has a correspondent in S_2 .
 - DEPENDENCE: Every element of S_2 has a correspondent in S_1 .
- (4) OO-correspondence (Benua 1997)
 - OO-FAITH-X: Every element of the derivative form has a correspondent in the base, and vice versa.
- (5) Anti-Faithfulness (Alderete 1999):
 - For every Faithfulness constraint F , there is a corresponding Anti-Faithfulness constraint $\neg F$ that is satisfied in a string S iff S has at least one violation of F .

A more detailed definition is given in (6):

- (6) Anti-Faithfulness constraints (Alderete 2000):
 - $\neg \text{MAX-X: } \neg [\forall x \exists x' [x \in S_1 \rightarrow x' \in S_2 \ \& \ xRx']]$
 ‘If there is one, delete (at least) one X in the $S_1 \rightarrow S_2$ mapping.’
 - $\neg \text{DEP-X: } \neg [\forall x \exists x' [x \in S_2 \rightarrow x' \in S_1 \ \& \ xRx']]$
 ‘Insert (at least) one X in S_2 not present in S_1 .’
 - $\neg \text{IDENT(F): } \neg [\forall y \forall y' \forall F [yRy' \rightarrow y =_F y']]$
 ‘(At least) one pair of correspondent segments must differ in feature F .’

Translating that into constraints restricting accent, $\neg \text{MAX-ACCENT}$ brings about an obligatory deletion of accent. $\neg \text{DEP-ACCENT}$ requires the insertion of an accent in the derivative form, where no accent was in the base (this accounts for pre- and postaccenting affixes); and last but not least, $\neg \text{IDENT(ACCENT)}$ (Alderete 1999, $\neg \text{NO-FLOP-PROM}$ in Alderete 2000) causes an obligatory accent shift, because the accent of the derivative form should be in another position than in the base. Ranked in the following way, the different accentual processes can be captured schematically (7).

- (7) Schematic rankings:
 - a) $\neg \text{OO-MAX-ACCENT} \gg \text{OO-MAX-ACCENT} \rightarrow \text{Accent Deletion}$
 - b) $\neg \text{OO-DEP-ACCENT} \gg \text{OO-DEP-ACCENT} \rightarrow \text{Pre-/Post-Accentuation}$
 - c) $\neg \text{OO-IDENT-ACCENT} \gg \text{OO-IDENT-ACCENT} \rightarrow \text{Accent Shift}$

3. Accentual processes involving AF

TAF was developed, among other things such as a reversal in voicing specification in Luo (Gregersen 1972, Alderete 1999, 2000), for accentual phenomena such as accent deletion, pre- and postaccentuation, and accent shift. These accentual processes are mainly due to interactions between morphemes. Morphologically triggered accent can be said to involve lexical markings in that morphemes can be specified for accent in the lexicon. Morphemes can thus be self-accenting (they carry an accent themselves), pre- and postaccenting (they assign accent on the preceding or following morpheme/syllable), or they can trigger an accent shift (the accent stays on the original

morpheme, but switches to another vocalic peak). Since it is generally the case that languages only realize one main accent per word, a conflict is predicted in situations where two inherently accented morphemes are combined to form a word. There are several ways to solve such a conflict. One possibility is to choose the accent that is close to a designated edge of a word (coded in the grammar in the form of alignment constraints, for instance). Another possibility is to give some of the morphemes a dominant status, in opposition to the others. The conflict is then solved in favour of the specification of the dominant morpheme. With TAF, these conflicts are solved in terms of anti-output-output correspondence. While a similarity between related forms is due to OO-correspondence, a dissimilarity between such forms is due to a principle that requires a difference to make clear the diverse morphological status of the forms. So, in derived environments, such as paradigms, where each entry might code additional information, a difference between forms can not only be expressed through addition of structure, but also through alternation of already existing structure (i.e. accent deletion or insertion). As a result, a distinction between root-controlled accent and affix-controlled accent emerges (Alderete 1999, 2000). In languages with root-controlled accent (i.e. Cupeño, Hill and Hill 1968), roots as well as affixes can possess a lexical specification for accent. If a specified root and a specified affix are combined, the root accent overrides the accent of the affix. These languages resemble the general tendency of a cross-linguistically observable ranking of RootFaith above AffixFaith (McCarthy & Prince 1995). In opposition to that are systems with affix-controlled accent, where affixes can be the dominant morphemes and trigger accent. So a reversal of AffixFaith over RootFaith can be observed. This reversal is accounted for by TAF in that it requires the opposite of Faithfulness.

To classify a system as root-controlled or affix-controlled, several criteria have to be fulfilled. In order to be classified as being root-controlled, a system has to be methodically faithful to the root properties. To be classified as affix-controlled, affix faithfulness has to outrank root faithfulness in at least some cases, that is, some affixes have to be dominant, while others can be recessive, in the sense that accent conflicts are solved in favour of dominant affixes, but not in favour of recessive ones. In affix-controlled systems, affix morphemes generally display more contrast in their specifications, in the sense that e.g. roots as well as affixes can be inherently accented, but that only affixes can be also pre- or postaccenting. In TAF, this is expressed by the principle of Strict Base Mutation (Alderete 1999, 2000), which requires that affix-controlled processes always affect an element of the base of a morphological process. So in affix-controlled systems, it should for instance not be the case that a root can override affix properties, in the sense that there are no pre- or postaccenting roots.

In the following section, I want to apply TAF to Modern Greek, a language whose accent system involves accentual conflicts resulting in accent deletion, pre- and postaccentuation. It will turn out that TAF cannot satisfyingly account for the attested accent pattern. But first, I will present some properties of Modern Greek.

4. TAF analysis of Modern Greek

4.1. Modern Greek – the data

Modern Greek is a language with fusional morphology, where affixation is the common process of word formation. Generally, a word consists of a root and an inflectional

suffix. Derivational suffixes can be inserted between root and inflection. These morphemes can be lexically marked with certain accent properties. Roots can be specified for accent, and for being postaccenting, that is posing the accent on the following morpheme. Inflectional suffixes can be specified for accent, and for being preaccenting. Derivational suffixes can be specified for accent, but also for pre- or postaccenting (in these classifications, I follow Revithiadou 1999). The language allows only one main accent per word, so the morphemes of a word can compete with each other regarding accent assignment. It appears that there seems to be a hierarchy: Derivational morpheme >> root morpheme >> inflectional morpheme. If only root and inflection are combined, it is always the root accent that appears on the surface (8a, c, f, g), except if the root is not inherently accented (8d, e). Only then can an inflectional affix maintain its lexical specification on the surface (8e). If the inflection is not inherently specified, a phonological default is assigned to the antepenultimate syllable (if the word contains three or more syllables, like in (8d); if there are less syllables, accent is assigned to the initial syllable). When a (lexically specified) derivational suffix is added, root as well as inflectional accent is overridden (8b). The affixes can thus be subdivided into dominant and recessive morphemes.

(8) Distribution of accent in Modern Greek

a) /stafíð/ _{root}	+ /a/ _{infl}	→ stafíða	‘grape-Nom-Sg’
b) /stafíð/ _{root}	+ /á/ _{deriv} + /i/ _{infl}	→ stafíðáki	‘grape-Nom-Sg-Dim’
c) /klívan/ _{root}	+ /os/ _{infl}	→ klívanos	‘kiln-Nom-Sg’
d) /anθrop/ _{root}	+ /os/ _{infl}	→ ánthropos	‘man-Nom-Sg’
e) /anθrop/ _{root}	+ /u/ _{infl}	→ anθrópu	‘man-Gen-Sg’
f) /uran/’ _{root}	+ /os/	→ uranós	‘sky-Nom-Sg’
g) /uran/’ _{root}	+ /u/ _{infl}	→ uranú	‘sky-Gen-Sg’

From these data, we can already note a problem with the typology TAF requires: Modern Greek does involve postaccenting roots (8f, g), that is, roots are able to assign an accent to the following morpheme. In the next section, I demonstrate how TAF can be applied to Modern Greek, and where it fails to account for the facts. To keep it short, I will limit my discussion on self-accented and postaccenting roots, dominant self-accented and recessive preaccenting suffixes. There are also accentless and self-accented recessive suffixes, as well as pre- and postaccenting dominant suffixes (Revithiadou 1999), to be complete, but the forms mentioned in (8) will be sufficient to explain the difficulty.

4.2. Modern Greek – root-controlled or affix-controlled?

Since the Modern Greek accent system employs both dominant and recessive affixes, the language can be classified as having affix-controlled accent. Besides, pre- and postaccentuation is also involved. So some of the ranking of (7) should be found in Modern Greek, namely \neg OO-MAX-ACCENT >> OO-MAX-ACCENT for the dominance effects and \neg OO-DEP-ACCENT >> OO-DEP-ACCENT for the occurrence of pre- and postaccentuation. To account for the fact that there are dominant and recessive affixes, the Anti-Faithfulness constraints have to be split into \neg OO_{Dom}-MAX-ACCENT and \neg OO_{Rec}-MAX-ACCENT, respectively. \neg OO_{Rec}-MAX-ACCENT is ranked below OO-MAX-ACCENT and IO-MAX-ACCENT, because recessive affixes do not override root accent.

Let us first consider the case in Modern Greek, when a root is combined with a dominant self-accented affix: /stafið-ák-i/. A candidate pair like in (9a) violates the Anti-Faithfulness constraint referring to the dominant morpheme -ak- (\neg OO_{Dom}-MAX-ACC), because the accent of the base is not deleted. A candidate pair like that in (9b), where no accent is realized, does not violate \neg OO_{Dom}-MAX-ACC, because this constraint requires a deletion of the accent in the base part, which is fulfilled in this candidate. But this candidate is worse than (9c), because it has one more violation of IO-MAX-ACCENT, so candidate (9c) surfaces as optimal.³

(9) *stafíða* - *stafíðaki*⁴

Base:	/stafíð/+/áki _{Dom} /	\neg OO _{Dom} - MAX-ACC	OO-MAX-ACC	IO-MAX-ACC
a. stafíða	stafíðaki	*!		*
b. stafíða	stafíðaki		*	**!
c. stafíða	stafíðaki		*	*

Anti-Faithfulness would even account for the case where a root and a recessive preaccenting inflectional suffix are combined (10): accent remains on the root (10d) because the added suffix is recessive. Since no dominant affixes are involved in (10), \neg OO_{Dom}-MAX-ACC is not violated at all, so it was spared from the tableau. The positive counterpart, OO-MAX-ACC, is violated in candidate pair (10b) as well as in pair (10c) because in both cases the accent of the base is deleted. Candidate (10b) looks like candidate (10a), but it gets its accent from the inflection. IO-MAX-ACC is violated in (10b+c), because the specification of the root has been deleted. Note that in candidate (10a), not the accent of the root is deleted, but rather shifted to the right. This is a violation of high-ranked IDENTITY (Beckmann 1997), which requires that correspondents are the same. The candidate pair (10d) on the other hand only violates the constraint referring to the recessive inflection. Remember that \neg OO_{Rec}-DEP-ACC is the constraint that causes an insertion of an accent on the base, and thus is responsible for preaccentuation.

(10) *klívanos* - *klívanu*

Base:	/klí ₁ van/+’ ₂ /u _{Rec} /	OO- IDENT	OO-MAX- ACC	IO-MAX- ACC	\neg OO _{Rec} - DEP-ACC
a. klívanos	klivá ₁ nu	*!			*
b. klívanos	klivá ₂ nu		*(!)	*(!)	
c. klívanos	klivanu		*(!)	*(!)	*
d. klívanos	klívanu				*

However, if you take the output-output pair *ánthropos* - *anθrópu* into consideration and compare it with the pair in (10), a formal problem arises. Since both *klívanos* and

³ In Modern Greek, there are also other constraints involved in accentuation. Nevertheless, the focus is here on the interaction between the constraints referring to dominant and recessive morphemes, and the conflict between lexical specifications of these morphemes. I will come back to the other constraints in my own analysis later on.

⁴ The suffix -aki- consists, strictly speaking, of the derivation -ak- and the inflection -i-, but since the inflection is no point of interest here, I contract both morphemes and treat them as a unit.

ánthropos serve as a base, they both have already been assigned accent. So you would expect the same results when the preaccenting suffix *-u* gets attached. But, in one case accent remains on the original position of the base (*klívanu*), while in the other the accent is realized in pre-position of the inflection (*ánthrópu*). According to TAF and the ranking established in (9) and (10), accent should be realized on the same position in the *ánthropos* - *ánthrópu* pair as in *klívanos* – *klívanu*, namely on the base. But this is not the case, as (11) shows. $\neg\text{OO}_{\text{Dom}}\text{-MAX-ACC}$ is not of interest here because no dominant affixes are involved, so it is left out of the tableau. The candidate pair (11c) is a loser, because it violates OO-MAX-ACCENT in the sense that the accent of the base is deleted and another accent is assigned elsewhere. This candidate also violates $\neg\text{OO}_{\text{Rec}}\text{-DEP-ACC}$, since it has an accent on the inflection, rather than being preaccenting. But this is not crucial here. The candidate pair in (11a) wins according to TAF because it has only one violation of low-ranked $\neg\text{OO}_{\text{Rec}}\text{-DEP-ACC}$, but instead, (11b) should win (because that is the attested form in Modern Greek), although it violates general OO-MAX-ACCENT , since the accent of the base is deleted and an accent is assigned by the inflection. Thus, according to Anti-Faithfulness, candidate (11a) should win, but in fact, candidate (11b) is attested in Modern Greek.

(11) *ánthropos* - *ánthrópu*

Base:	/ánthrop/+’/u _{Rec} /	OO-MAX-ACC	IO-MAX-ACC	$\neg\text{OO}_{\text{Rec}}\text{-DEP-ACC}$
⊗ a. <i>ánthropos</i>	<i>ánthropu</i>			*
● [*] b. <i>ánthropos</i>	<i>ánthrópu</i>	*!		
c. <i>ánthropos</i>	<i>ánthropú</i>	*!		*

One could think of a repair strategy, specifically to subdivide OO-MAX-ACC , which requires faithfulness to the base, into a dominant and a recessive constraint, so that the dominant constraint is fulfilled in a case like (10), and that the recessive version of this constraint is ranked below $\neg\text{OO}_{\text{Rec}}\text{-DEP-ACC}$, so that it is fulfilled in a case like (11). Yet, this would be contradictory to the prediction of TAF, namely Strict Base Mutation, which says that in an output-output relation only the base can be affected. The principle of SBM derived out of the observation that affixes show more variation in their properties than roots in affix-controlled systems. The affixes can not only be \pm accented, but also \pm dominant and also pre- and postaccenting. Roots on the other hand show according to Alderete (1999, 2000) cross-linguistically only up as \pm accented. One conclusion out of that is, if roots assign accent, it is not due to dominance, but due to the systematic emergence of $\text{RootFaith} \gg \text{AffixFaith}$. Roots might then realize their inherent specification only when combined with a recessive affix. In opposition to that stands Modern Greek, where you would have to distinguish between dominant and recessive roots, where $\text{Affix}_{\text{Dom}} \gg \text{Root}_{\text{Dom}} \gg \text{Affix}_{\text{Rec}} \gg \text{Root}_{\text{Rec}}$.

The reason why forms like *ánthrópu* and *klívanu* behave differently is that they are accented on the antepenultimate syllable due to different principles. A root like *klivan-* is lexically specified for accent, so in cases where lexically specified morphemes are involved, IO-Faithfulness is responsible for accent assignment. A root like *ánthrop-* is lexically unmarked and thus does not fall under the scope of IO-Faithfulness . Accent in *ánthropos* is assigned to the antepenultimate syllable by way of a phonological default (Malikouti-Drachman & Drachman 1989, Revithiadou 1999). Anti-Faithfulness constraints have no impact on this difference in roots, since they only compare outputs with each other, and have no impact on input structure.

Furthermore, Modern Greek has poststressing roots (8f, g), similar to the ones observable in Russian. While postaccenting roots in Russian were analysed by Alderete as not having a lexical specification, but being the default due to (a positive Faithfulness) constraint POST-STEM-ACCENT (Alderete 2000), this analysis cannot be transferred to Modern Greek. Post-root accent is definitely not the default accent position in the language. The default position is the antepenultimate syllable (if a word is three or more syllables long, else accent is placed on the initial syllable; Malikouti-Drachman & Drachman 1989, see Revithiadou 1999 for an OT analysis), which in OT terms can be expressed with constraints like NONFINALITY and constraints referring to foot structure. Postaccenting roots must have a lexical specification for accent, else they could not assign stress to the succeeding morpheme (which they clearly do in cases like 8f and g).

Therefore, I would like to propose another approach to the Greek data. The conflict between the lexical specifications of roots and affixes can be solved without OO-Faithfulness, if one assumes that the lexical accent is coded as partial foot structure (Revithiadou 1999, Apoussidou 2002; Inkelas 1998 on full foot structure as lexical specifications). In that case, IO-Faithfulness constraints referring to the strong part of a foot (self-accenting and postaccenting morphemes) or to the weak part of a foot (preaccenting morphemes) are required.

I will sketch this approach shortly below⁵, starting with the default accent of a word like *ánthropos*. In this case, no lexical specification is involved. Constraints involved are TROCHEE (since Modern Greek is a trochaic language), FOOTBINARITY, NONFINALITY (in the sense of Tesar and Smolensky, 2000, that is the final syllable of a word should not be footed; this constraint is responsible for extrametricality), and ALIGN-FOOT-RIGHT (that prevents four or more syllable words from being accented on the initial syllable). The candidates (12a) and (12b) are both ruled out, because they both violate the constraint responsible for extrametricality; also, candidate (12b) violates FTBINARITY. Candidate (12a) is the winner, because it only violates low-ranked ALIGN-Ft-R.

(12) <i>ánthropos</i>		TROCHEE	FTBIN	NONFIN	ALIGN-Ft-R
/anθrop/+/os/					
☞ a. (ánthro)pos					*
b. an(θrópos)				*!	
c. anθro(pós)			*(!)	*(!)	

As soon as at least one lexical specification for accent is involved, the phonological default is overridden. Thus, the constraints referring to the lexical specifications are higher ranked than the constraints that assign the default accent. This is demonstrated in (13). A constraint MAX-FOOT_{weak} requires the lexical specification in form of a weak part of a foot to surface (in this case, the specification of the inflectional suffix). IDENTITY requires the lexical specification to surface on the morpheme that is marked for it. Candidate (13a) now violates high-ranked IDENTITY, since the specification of the suffix is realized elsewhere in the word, and not on the inflection itself. Candidate (13c) violates MAX-FOOT_{weak} because the specification does not surface as weak, but on the contrary is realized as strong, that is, the preaccenting suffixes is realized as accented. Candidate (13b) is optimal, since it violates neither of the higher ranked constraints.

⁵ See Apoussidou 2002 for a more elaborate analysis of Modern Greek accent.

(13) *anθrópu*

/anθrop/+u/	MAX-FOOT _{strong}	IO-IDENTITY	FTBIN	NONFIN
a. (ánθro)pu		*!		
☞ b. an(θrópu)				*
c. anθro(pú)	*!		*	*

Similarly, if a morpheme is specified for a strong part of a foot, a constraint MAX-FOOT_{strong} requires that the specification surfaces (14). A candidate that contains an accent shift is ruled out by IDENTITY (14b), while a candidate without any accent would violate MAX-Ft_{strong} (14c). A candidate like (14a) is optimal, since it is faithful in any respect.

(14) *klívanos*

/(klivan/+os/	MAX-Ft _{strong}	IO-IDENTITY	FTBIN	NONFIN
☞ a. (klíva)nos				
b. kli(vános)		*!		*
c. klivanos	*!			

In (15), a conflict between two specifications is shown. It can be solved through establishing a ranking order between the two faithfulness constraints MAX-Ft_{strong} and MAX-Ft_{weak}, so the candidate that realizes the strong specification of a foot (15a) wins. A candidate like (15c) is suboptimal, since it violates both faithfulness to the specification as well as to the position of the specification.

(15) *klívanu*

/(klivan/+u/	MAX-Ft _{strong}	MAX-Ft _{weak}	IO-IDENTITY	FTBIN	NONFIN
☞ a. (klíva)nu		*			
b. kli(vánu)	*!				*
c. kliva(nú)		*	*!*	*	*

Postaccenting roots can be regarded as having a specification for a strong part of a foot as well, and this specification is linked to the final segment of the root. Then, the analysis looks like the ones for words like *klívanos* and *klívanu*. Still, in cases where an accented derivational suffix is added, a distinction has to be made between dominant and recessive faithfulness (Revithiadou 1999), since then neither faithfulness to the kind of specification nor to the position of the specification can solve the conflict.

Furthermore, one could argue that Russian has also postaccenting roots in form of a lexical specification. If one assumes that the default accent is on the leftmost syllable in a word (like e.g. Revithiadou, 1999, does), then postaccenting roots in Russian would pose the same problems to TAF as Modern Greek does.

5. Conclusion

While TAF might be applied quite successfully to featural changes in OO-relations (values for the feature voice in Luo, Alderete 1999, 2000; Turkish reduplication, Kelepir 1999), it is not the adequate instrument to account for accent phenomena, since it cannot distinguish between differences in bases. Also, there are some formal

problems: how can a constraint, which is sensitive to output-output relations, have access to lexical specifications of morphemes? Lexical specifications are only accessible via input-output faithfulness.

Other theoretical approaches might deal with the same issues in a somehow better way, for instance Realizational Morphology Theory (RMT, Kurisu 2001), where differences in morphologically related forms is due to addition of information (which may not necessarily be realised in addition of structure, but also in distraction of structure). Questions remain to what extent attempts to restrict Anti-Faithfulness to morphology (i.e. Horwood 2002) could improve this approach.

References

- Alderete J 1999 Morphologically Governed Accent in Optimality Theory. Doctoral Dissertation (Amherst: Univ. of Massachusetts)
- 2000 Dominance Effects as Transderivational Anti-Faithfulness. ROA Nr 407
- Apoussidou D 2002 Unpredictable Accent Patterns in Correspondence Theory. SFB-282-Paper Nr. 120 (Düsseldorf: Heinrich-Heine-Univ.)
- Benua L 1997 Transderivational Identity: Phonological Relations between Words. Doctoral Dissertation (Amherst: Univ. of Massachusetts)
- Gregersen E A 1972 Consonant Polarity in Nilotic. In E Voeltz (ed) Third Ann. Conf. on African Linguistics (Bloomington: Indiana Univ.)
- Hill J and Hill K 1968 Stress in the Cupan (Uto-Aztecan) Languages. Intern. Journal of American Linguistics 34 p 233-241
- Horwood G 2002 Anti-Faithfulness and Subtractive Morphology. ROA Nr 466
- Inkelas S 1998 Exceptional stress-attracting suffixes in Turkish: representations versus the grammar. In The Prosody-Morphology Interface
- Kelepir M 1999 Turkish emphatic reduplication and antifaithfulness. Proc. ConSOLE VII p 153-167
- Kurisu K 2001 The phonology of morpheme realization. Doctoral Dissertation (Santa Cruz: Univ. of California)
- Malikouti-Drachman A and Drachman G 1989 Stress in Greek [Tonismos sta Ellinika]. Studies in Greek Linguistics 1989 (Univ. of Thessaloniki) p 127-143
- McCarthy J and Prince A 1993 Prosodic Morphology I: Constraint Interaction and Satisfaction. Rutgers Center for Cognitive Science (RuCCS) Technical Report 3 (Amherst: Univ. of Massachusetts, and Rutgers Univ.)
- McCarthy J and Prince A 1995 Faithfulness and Reduplicative Identity. ROA Nr 60-0000
- Prince A and Smolensky P 1993 Optimality Theory: Constraint interaction in Generative Grammar. Ms. (Rutgers Univ. and Univ. of Colorado at Boulder)
- Revithiadou A 1999 Headmost Accent Wins. Head Dominance and Ideal Prosodic Form in Lexical Accent Systems. The Hague: Holland Academic Graphics
- Tesar B and Smolensky P 2000 Learnability in Optimality Theory (Cambridge, London: MIT Press)
- Beckmann J 1997 Positional Faithfulness. Doct. Dissertation (Amherst: Univ. of Massachusetts)

The acquisition of phonological opacity

Ricardo Bermúdez-Otero

University of Newcastle upon Tyne

Abstract. This paper argues that Stratal OT is explanatorily superior to alternative OT treatments of phonological opacity (notably, Sympathy Theory). It shows that Stratal OT supports a learning model that accounts for the acquisition of opaque grammars with a minimum of machinery. The model is illustrated with a case study of the classic counterbleeding interaction between Diphthong Raising and Flapping in Canadian English.

1. Phonological opacity: Stratal OT vs Sympathy Theory

Following the appearance of Prince & Smolensky (1993), phonologists were quick to realize that, in its original version, OT was unable to describe a large set of phonological phenomena previously modelled by means of opaque rules. Ten years later, opacity remains the severest challenge confronting OT phonology. The problem is crucial because opacity effects constitute one of the clearest instances of Plato's Problem in phonology: learners face the task of acquiring generalizations that are not true on the surface. The ability to explain the acquisition of opaque grammars should accordingly be regarded as one of the main criteria by which generative theories of phonology are to be judged. Among the variants of OT currently on offer, two claim to provide a comprehensive solution to the problem of opacity: Sympathy Theory (McCarthy 1999, 2003) and Stratal OT (Bermúdez-Otero 1999; Kiparsky 2000). This paper compares the strategies whereby these two phonological models seek to achieve explanatory adequacy.

2. Weak explanatory adequacy: typological restrictiveness

A theory of grammar is said to attain 'explanatory adequacy' when it solves the logical problem of language acquisition. However, the term is often used in a watered-down sense equivalent to 'typological restrictiveness': on the assumption that learnability improves in proportion with reductions in the size of the grammar space generated by UG, grammatical frameworks that are typologically restrictive are often felt to be more explanatory (but cf. §3 below). In this section, therefore, I look at the space of possible opacity effects defined by Stratal OT (§2.1) and by Sympathy Theory (§2.2).

2.1. Stratal OT

Stratal OT borrows two key ideas from previous generative theories of phonology. The first is the *phonological cycle*. In a cyclic framework, given a linguistic expression e with a phonological input representation I , the phonological function P applies recursively from the inside out within a nested hierarchy of phonological domains associated with (but not necessarily fully isomorphic with) the morphosyntactic constituent structure of e : i.e. if $I = [[x][[y]z]]$, then $P(I) = P(P(x), P(P(y), z))$. The second key idea is *level*

segregation, according to which the phonology of a language does not consist of a single function P , but of a set of distinct functions or ‘cophonologies’ $\{P_1, P_2, \dots, P_n\}$, such that the specific function P_i applying to domains of type δ_i is determined by the type of morphosyntactic construction associated with δ_i (e.g. a stem, word, or phrase).

In Stratal OT, therefore, opacity arises from the serial interaction between cycles. Within each cycle, however, the input-output mapping is transparent, as it is effected in the parallel fashion that characterizes classical OT. In other words, each cycle involves a single pass through *Gen* and *Eval*: i.e. $P_i(\delta_i) = Eval_i(Gen(\delta_i))$. This principle imposes severe restrictions on the complexity of opaque interactions. Notably, the depth of derivations is bound by the number of cycles, which is in turn independently constrained by the morphosyntactic structure of the linguistic expression. In addition, the phonology of the most inclusive domain (corresponding to processes applying across the board at the level of the Phonological Utterance) is predicted to be transparent.

2.2. *Sympathy Theory*

In Sympathy Theory, apparent misapplication is caused by a set of constraints, called ‘sympathy constraints’, which enforce identity between the output and a failed co-candidate endowed with special status: the ‘sympathetic candidate’ (or ‘ \otimes -candidate’). This candidate is defined as the most harmonic among the subset of candidates satisfying a designated ‘selector constraint’ (or ‘ \star -constraint’).

The theory, however, requires a number of additional stipulations. Unlike input-output faithfulness, for example, sympathy must be an asymmetric relationship: the output can copy properties of the \otimes -candidate but not vice versa, for otherwise opaque underapplication would be impossible (Bermúdez-Otero 1999: 143-148). In contrast, IO-correspondence is symmetrical and reversible: outputs are faithful to the corresponding inputs in production, whereas in acquisition inputs are modelled upon outputs by Input Optimization (see §4.4 below). McCarthy (1999: 339) secures the asymmetry of sympathetic correspondence by means of the following stipulation:

(1) **Invisibility of sympathy constraints**

Selection of sympathetic candidates is done without reference to sympathy constraints.

Interestingly, this proviso imposes a significant restriction upon opacity effects. When two or more sympathetic candidates are active in a single computation, each is selected independently and affects the evaluation of output candidates in parallel. Sympathy Theory can therefore mimic serial derivations that involve at most one intermediate step. Significantly, this empirical prediction turns out to be false: Bermúdez-Otero (2002) adduces a counterexample from Catalan where two intermediate representations are crucially needed.

Further to constrain the generative power of sympathy, McCarthy (1999: 339) adds another principle to the theory:

(2) **\star -confinement**

The selection of a sympathetic candidate must be confined to a subset of candidates that obey an IO-faithfulness constraint F .

This stipulation reduces the number of possible selector constraints and, therefore, the number of possible sympathetic relationships. In addition, it enables McCarthy to rationalize sympathy as a kind of ‘faithfulness by proxy’, where the optimal output

copies some property of a hyperfaithful failed co-candidate. Empirically, however, the principle of \star -confinement has been shown to cause undergeneration (Itô & Mester 1997; de Lacy 1998; Bermúdez-Otero 1999: 150-191). The characterization of sympathy as ‘faithfulness by proxy’, moreover, does not translate into functional gains in terms of improved lexical access, for, as McCarthy (1999: 343) himself acknowledges, opaque processes are often neutralizing (see Bermúdez-Otero 1999: 152).

In a final bid to restrict the complexity of sympathetic effects, McCarthy has also adopted special measures against non-paradigmatic non-vacuous Duke-of-York gambits. In serial terms, a Duke-of-York derivation has the form $a \rightarrow (\dots) \rightarrow b \rightarrow (\dots) \rightarrow a$; it is non-vacuous if b either escapes a process applicable to a (‘bleeding’) or undergoes a process not applicable to a (‘feeding’); it is non-paradigmatic if b does not surface as (part of) a grammatically related expression. McCarthy claims that such derivations do not occur in natural language. To prevent Sympathy Theory from mimicking them, he resorts to a combination of two devices: one is the \star -confinement clause stated in (2); the other is an *ad hoc* principle of ‘cumulativity’ (McCarthy 1999: §4.2; 2003), which penalizes output candidates that are more faithful to the input than the \otimes -candidate.

Cumulativity is deeply problematic. First, it is simply false that non-paradigmatic non-vacuous Duke-of-York gambits do not occur in natural language. Such derivations do exist, and they are not hard to acquire provided that the phonological processes involved produce robust alternations (see §4 and §5 below); one such case is found in Catalan (Bermúdez-Otero 2002). Secondly, the formal stipulations to which McCarthy resorts are fraught with difficulties. As we have seen, \star -confinement is empirically untenable. In addition, it is only by brute force that the principle of cumulativity manages to block nonvacuous Duke-of-York gambits. Conceptually, moreover, cumulativity conflicts with the rationalization of sympathy as faithfulness by proxy.

3. Strong explanatory adequacy: the logical problem of language acquisition

As we have seen, Sympathy Theory fails in its attempts to define a highly restricted space of possible opacity effects. However, even if the theory attained this goal, the fact would be far less significant than McCarthy implies. This is because, in practice, typological restrictiveness does not guarantee explanatory adequacy in the strong sense. To appreciate this point, consider two theories of grammar T_1 and T_2 , which define the grammar spaces S_1 and S_2 respectively. If both S_1 and S_2 are too large for convergence to be guaranteed by brute-force searching, then the prime determinant of learnability will be the relative efficiency of the learning algorithms associated with T_1 and T_2 , rather than the relative size of S_1 and S_2 (see Tesar & Smolensky 2000: 2-3). In other words, a phonological model cannot achieve explanatory adequacy in respect of opacity simply by restricting the space of possible opaque effects; one must show that the learner is able to search that space effectively. Tellingly, there is to date no theory of the acquisition of sympathy-theoretic grammars (see McCarthy 1999: 340, note 9). In contrast, Stratal OT offers a straightforward recipe for the acquisition of opacity effects (§4-§5).

4. Phonological acquisition in Stratal OT: overview

This section presents the key ingredients for a model of phonological acquisition in Stratal OT. As we shall see in §5, this model effectively accounts for the acquisition of opacity effects supported by evidence from alternations. The model achieves this by making the most of the assets of the synchronic theory: notably, it fully exploits the serial interaction between strata and the intimate connection between the morphosyntactic domain of a phonological process and its stratal ascription (§4.1, §4.2). Beyond this, the model simply adopts current solutions to the problem of acquiring constraint rankings

and input representations (§4.3, §4.4): the only provision added specifically to deal with opaque phenomena is the principle of Archiphonemic Prudence (§4.5).

4.1. *Iterative stratum construction*

Stratal OT enables one to break the logical problem of phonological acquisition down into a set of relatively simpler subproblems, for learning a phonological grammar consists of acquiring a series of cophonologies: typically, the phrase-level, word-level, and stem-level cophonologies. Moreover, since the input to level n provides the output of level $n-1$, each of these subproblems can be tackled in a logical progression. Acquiring the phrase-level cophonology, for example, involves (i) discovering the phrase-level constraint hierarchy and (ii) assigning single representations to individual words at the input to the phrase level. The input representations so assigned constitute the output of the word level and provide the data for the next iteration in the process of acquisition.

4.2. *The emergence of opacity*

As we saw in §2.1, opacity arises from interactions between processes that apply transparently in their own strata: each phonological generalization in the grammar holds true in the output of the corresponding level, which defines the domain of the generalization. During acquisition, therefore, the task of assigning phonological processes to the appropriate strata can be reduced to the independent problem of discovering correct input representations.¹ Consider, for example, a process p that applies at level n and is rendered opaque by changes introduced at level $n+1$. If input representations are correctly assigned at level $n+1$, p will be true of the output of n . On this basis, any of the standard ranking algorithms designed to acquire transparent processes will establish the ranking for p in the constraint hierarchy of n . By the same token, the constraint ranking for p will not be introduced at level $n+1$ simply because the ranking algorithm encounters contradictory data, as p does not hold true in the output of $n+1$. In other words, the learning model need do no more than establish transparent constraint rankings (§4.3) and assign input representations correctly (§4.4, §4.5); the grammatical architecture of Stratal OT takes care of the rest.

4.3. *Constraint ranking by pure phonotactic learning under the identity map*

At any level, then, the first task for the learner is to find the appropriate ranking of constraints, given a set of output forms. As Prince & Tesar (1999) and Hayes (1999) have shown, this can be done largely on the basis of purely distributional information: assuming the identity map (input = output) plus a MARKEDNESS » FAITHFULNESS bias (henceforth, ‘M » F bias’), the learner must demote markedness constraints and promote faithfulness constraints just enough to derive the output from identical input. The details of the ranking algorithm need not concern us here. The important point, rather, is that alternations usually conspire to bring morphological or syntactic collocations in line with output phonotactics; for this reason, pure phonotactic learning will in most cases suffice to find the constraint rankings driving not only phonotactics but also alternations. The acquisition of the latter will then boil down to mere input assignment (Hayes 1999: §6).²

¹ In Stratal OT, only the highest grammatical level is subject to Richness of the Base. The input to a non-initial stratum n will possess systematic properties enforced by the constraint hierarchy of level $n-1$.

² In some cases, constraints can be ranked appropriately only if the correct input representations are known. This problem can usually be solved by iterating between constraint ranking and input assignment until equilibrium is reached (see Tesar & Smolensky 2000: §1.3.2, §5.2).

4.4. *Input assignment (I): alternations prompt departures from the identity map*

Following the currently prevalent view, I assume that learners need evidence from alternations in order to depart from the identity map. In line with the principle of Input Optimization (Prince & Smolensky 1993: §9.3),³ departures from the identity map are minimal: unwarranted disparity between inputs and outputs causes unnecessary violations of faithfulness constraints. Unfortunately, currently available formulations of Input Optimization for alternating items (e.g. Inkelas 1995) are flawed. Bermúdez-Otero (in preparation) develops an alternative supported with diachronic evidence from changes involving input restructuring. This can be summarized as follows:

(3) **Input optimality (after Bermúdez-Otero in preparation)**

An input representation is optimal iff it has no competitor that

- generates an identical set of output alternants,
 - generates all output alternants no less efficiently,
- and
- generates some output alternant more efficiently.⁴

In practice, this definition of input optimality selects a set J of potential inputs whose members are all output-equivalent and where each member is maximally similar to some output alternant. If the cardinality of J is greater than 1, the learner can make a (provisional) choice among its members by means of certain heuristics:

(4) a. **Hale's heuristic (after Hale 1973: 420)**

Prefer inputs that are well-formed outputs.

b. **Heuristic for asymmetric paradigms**

In an asymmetric paradigm, prefer those inputs which generate the central member of the paradigm most efficiently.

In (4b), the term ‘paradigm asymmetry’ refers to the well-known observation that citation forms often enjoy a special status in comparison with sandhi forms, that the nominative singular may be more central than other members of nominal paradigms, and so forth.

4.5. *Input assignment (II): Archiphonemic Prudence*

The final task for the learner is to assign input representations to non-alternating items. At this point, it is essential for the acquisition of opaque grammars that the learner should be able to use evidence from alternations to detect deviations from the identity map in non-alternating items. I suggest that this can be achieved by supplementing current learning models with a principle of ‘Archiphonemic Prudence’, designed to deal with possible instances of neutralization in non-alternating environments.

Let there be two input elements $/\alpha/$ and $/\beta/$ at level n , such that, in the output of n , the contrast between $/\alpha/$ and $/\beta/$ is maintained in environment $[__]_e$ and neutralized in environment $[__]_f$. Let γ be the output realization of $/\alpha/$ and $/\beta/$ in the neutralizing environment $[__]_f$. In such circumstances, the output of n will contain alternations such as $[\alpha]_e \sim [\gamma]_f$ and $[\beta]_e \sim [\gamma]_f$. We may refer to any token of $[\gamma]_f$ in the output of n as an

³ The original term ‘Lexicon Optimization’ is inappropriate in Stratal OT, where unmotivated disparity between inputs and outputs is avoided in all strata but inputs coincide with underlying representations only at the highest level.

⁴ More efficient inputs cause fewer violations of high-ranking faithfulness constraints. Note that input choice can only affect faithfulness, as markedness constraints only evaluate output forms.

‘archiphonemic string’. The problem arises when the learner comes across such an archiphonemic string in a non-alternating item *i*.⁵

I propose that, under Archiphonemic Prudence, the learner relies on the evidence from alternations such as $[\alpha]_e \sim [\gamma]_f$ and $[\beta]_e \sim [\gamma]_f$ to assign an input representation to *i* at level *n*. First, the learner creates two potential representations for *i* in the input to *n*: one where the input correspondent of γ is / α /, and another where the input correspondent of γ is / β /. The input candidates are otherwise identical with the output realization of *i* (recall that deviations from the identity map are minimal). These input candidates are then ‘quarantined’: they are not included in the data set triggering phonological acquisition at level *n*–1; learning at *n*–1 proceeds exclusively on the basis of non-quarantined inputs to *n*. When the constraint hierarchy of level *n*–1 is known, the learner is in a position to choose between the two quarantined candidates for input representation of *i* at level *n*: if the input candidate containing / α / is not a well-formed output at level *n*–1, the learner chooses the input candidate containing / β /. If both candidates are possible outputs at level *n*–1, they remain quarantined and the choice is passed on to level *n*–2.

5. Case study: Diphthong Raising and Flapping in Canadian English

In this section, the learning model outlined in §4 is applied to a classic empirical problem from Canadian English: the opaque interaction whereby the Flapping of /t/ (which also applies to /d/) counterbleeds the Raising of /aɪ/ and /aʊ/ to [əɪ] and [ʌʊ] before voiceless obstruents. As is well-known, this counterbleeding effect results in the apparent overapplication of Raising on the surface.⁶

(5)		<i>writing</i>	<i>riding</i>	<i>mitre</i>	<i>powder</i>
	UR	/raɪt-ɪŋ/	/raɪd-ɪŋ/	/maɪtər/	/paʊdər/
	Raising	rəɪtɪŋ	—	məɪtər	—
	Flapping	rəɪrɪŋ	raɪrɪŋ	məɪrər	paʊrər

Accounting for the acquisition of this opaque interaction is a highly significant result. Since it was first highlighted by Joos (1942), the problem has figured prominently in the theoretical debate (e.g. Chomsky 1964: 74). Kenstowicz (1994: 6-7) discusses it as a canonical example of Plato’s Problem in phonology and, significantly, Hayes (1999: §8) uses it to illustrate the challenges of learning morphophonological alternations in OT.

5.1. The target grammar

For the sake of concreteness, I assume foot-based analyses for both Flapping and Diphthong Raising (Jensen 2000). This choice, however, is irrelevant to the application of the learning model, which would operate in exactly the same way under an analysis based on ambisyllabicity.

Flapping involves the realization of /t/ and /d/ as [r] when (i) lax, (ii) preceded by a vowel or [r], and (iii) followed by a vowel. I assume, following Jensen (2000), that /t/ and /d/ are tensed at the word level if foot-initial; otherwise, they are lax (and so extra-short). Crucially for our purposes, Flapping is phrase-level, as indicated by the fact that it applies when its environment straddles a word boundary, as in (6c) and (6d):

⁵ As we shall see in §5.2, the learner can identify archiphonemic strings by examining sets of output alternants and factoring out the portions shared by all the members of each set.

⁶ In transcriptions, I ignore all allophonic detail not directly relevant to the discussion. In my choice of symbols for the diphthongs, I follow Wells (1982: §6.2.4). I am deeply grateful to my colleague Dr John Stonham for acting as a native speaker informant and for discussing with me the analysis presented in §5.1.

- | | | | | | | |
|-----|----|-------------|-------------------|-----|-------|------------|
| (6) | a. | [færər] | <i>fatter</i> | cf. | [fæt] | <i>fat</i> |
| | b. | [mærər] | <i>madder</i> | cf. | [mæd] | <i>mad</i> |
| | c. | [hi hɪr æn] | <i>he hit Ann</i> | cf. | [hit] | <i>hit</i> |
| | d. | [hi hɪr æn] | <i>he hid Ann</i> | cf. | [hid] | <i>hid</i> |

In the sentence given in (6c), the /t/ of *hit* is lax because it is not foot-initial at the word level; the /t/ only becomes prevocalic (and, in this case, also foot-initial by resyllabification) at the phrase level, where the words in the sentence are concatenated.

The diphthongs /aɪ/ and /aʊ/ undergo Raising to [əi] and [ʌʊ] when followed by a voiceless obstruent in the same foot.⁷ The examples in (7a) illustrate the rôle of consonant voicing; those in (12b), the rôle of foot structure.

- | | | | | | | |
|-----|----|-----------|---------------|-----|-------------|-----------------|
| (7) | a. | [nəɪf] | <i>knife</i> | cf. | [naɪvz] | <i>knives</i> |
| | | [hʌʊs] | <i>house</i> | cf. | [hʌʊzɪz] | <i>houses</i> |
| | b. | [ˈsəɪfən] | <i>syphon</i> | cf. | [saɪˈfənɪk] | <i>syphonic</i> |
| | | [səɪt] | <i>cite</i> | cf. | [saɪˈteɪʃn] | <i>citation</i> |

I suspect that, historically, Raising arose through the phonologization of a qualitative side effect of ‘Pre-Fortis Clipping’ (the shortening of vowels before fortis obstruents). Informally, I assume that the constraint hierarchy for Raising includes a context-free markedness constraint CLEARDIPH, which favours diphthongs where the auditory distance between the two elements is maximal; this constraint penalizes [əi] and [ʌʊ]. In the environment of Pre-Fortis Clipping, however, the context-sensitive markedness constraint CLIPDIPH demands that the distance between diphthongal elements should be minimized, thereby penalizing [aɪ] and [aʊ]. To be active, CLEARDIPH must dominate its faithfulness antagonist IDENT[mid], whilst CLIPDIPH must dominate IDENT[low]:⁸

- (8) a. **IDENT[mid]**
 Let α be an input segment, and let β be its output correspondent;
 if α is [mid], then β is [mid].
- b. **IDENT[low]**
 Let α be an input segment, and let β be its output correspondent;
 if α is [low], then β is [low].

Finally, the context-sensitive markedness constraint must dominate its context-free counterpart. Thus, the normal application of Raising requires the following rankings: CLEARDIPH » IDENT[mid], CLIPDIPH » IDENT[low], and CLIPDIPH » CLEARDIPH.

Crucially, there is clear evidence that Raising is ‘lexical’ (i.e. not phrase-level), as diphthongs are not raised when a voiceless obstruent follows across a word boundary:

- | | | | |
|-----|-----|---------------|---|
| (9) | | [ˈlaɪ fər mi] | <i>lie for me</i> |
| | cf. | [ˈləɪfər] | <i>lifer</i> (i.e. ‘convict serving a life sentence’) |

⁷ For our purposes, we could just as well assume an analysis where underlying /əi/ and /ʌʊ/ undergo lowering to [aɪ] and [aʊ] everywhere except before voiceless obstruents in the same foot; for our learning model, the choice is immaterial (see notes 1 and 12).

⁸ An analysis based on a symmetrical constraint IDENT[±low] would require learners to follow a slightly different learning path to that described in §5.2-§5.4, but would not be an obstacle to convergence.

In fact, Raising probably applies at the stem level. First, word-level suffixes such as *-ful* and *-ship* do not trigger Raising:

- (10) ['aɪfʊl] *eyeful*⁹ cf. ['əɪfəl] *Eiffel (Tower)*
 ['frʌʊʃɪp], *['frʌʊʃɪp] *Frauship* (nonce word derived from German
 Frau on the analogy of *lordship*, *ladyship*)

Secondly, Raising has lexical exceptions for some speakers (Wells 1982: 495): e.g. ['saɪkləps] *Cyclops* vs ['mæɪkrən] *micron*. Such behaviour is most often observed among phonological processes applying at the highest level in the grammar. Finally, Structure Preservation plays no rôle in Stratal OT (see Bermúdez-Otero 1999: 124) and so cannot be invoked as an argument against locating Raising in the stem level.

In sum, Diphthong Raising applies to stem domains, whereas the domain of Flapping is phrasal. From this information, Stratal OT correctly derives their relative order of application: phrase-level Flapping must follow (and so counterbleed) stem-level Raising.

How, then, can this system be acquired using the learning model described in §4? Setting up the constraint hierarchy for Flapping at the phrase level is clearly the easiest task: since Flapping is surface-true, the learner can achieve this by pure phonotactic learning from the primary data. In the case of Raising, in contrast, instances of surface overapplication (e.g. *writing*, *mitre*) and underapplication (e.g. *eyeful*, *lie for me*) will prevent the learner from establishing a raising hierarchy at the phrase level. Next, the learner must use the evidence from phrasal alternations such as *hit* vs *hit Ann* and *hid* vs *hid Ann* to discover the fact that surface [ɹ] derives from either /t/ or /d/ in the output of the word level, but —crucially— not from */ɹ/. In addition, the learner must be able to capitalize on this information and, using Archiphonemic Prudence, avoid the incorrect identity map */ɹ/→[ɹ] in non-alternating items such as /maɪtər/→[mɔɪrər] *mitre* and /vaɪtəl/→[vɔɪrəl] *vital*. If the learner chooses the correct input representations for alternating items at the phrase and word levels, Raising will become output-true at the stem level, and the learner will be able to establish the constraint ranking for Raising in the stem-level hierarchy by pure phonotactic learning. At this point, the learner can turn to items such as *mitre* and *vital*, previously quarantined under Archiphonemic Prudence. Since the stem-level constraint hierarchy enforces normal application of Raising, the incorrect phrase- and word-level inputs */mɔɪdər/ and */vɔɪdəl/ can be discarded, as they are ill-formed stem-level outputs. This just leaves the target input representations with /t/.

The success of this account rests upon two simple ideas. First, the constraint ranking driving a process *p* is established in the hierarchy of level *n* if and when *p* is true in the output of *n*; thus, the contrast between normal application and misapplication enables learners to assign phonological processes to the correct strata (§4.2). Secondly, learners depend on alternations to depart from the identity map either directly (in the case of alternating items; §4.4) or indirectly (when required by Archiphonemic Prudence; §4.5).

5.2. Acquiring the phrase-level cophology

If we ignore the problem of covert structure (see e.g. Tesar & Smolensky 2000: 6ff.), the primary linguistic data provide the child with direct access to the phrase-level output.

⁹ In this example, Raising is unlikely to be blocked by a weak foot over *-ful*. The word seems to be metrically equivalent to the unverbated compound ['həɪskʊl] *high school*, where Raising does apply (see Wells 1982: 494); cf. the unfused variant ['haɪ ,skʊl] .

Applying pure phonotactic learning to these data, the child will be able to establish the ranking for Flapping in the phrase-level constraint hierarchy, as Flapping is surface-true. In contrast, table (11) shows how the surface misapplication of Diphthong Raising prevents the learner from establishing the rankings CLEARDIPH » IDENT[mid] and CLIPDIPH » CLEARDIPH, which, as we saw in §5.1, are essential to the process.

(11)

<i>Datum</i>	<i>Triggered ranking</i>
məɪrər › məɪrər ‘mitre’	IDENT[mid] » CLEARDIPH
rəɪrɪŋ › rəɪrɪŋ ‘writing’	
aɪfʊl › əɪfʊl ‘eyeful’	CLEARDIPH » CLIPDIPH ¹⁰
lɑɪ fər mi › ləɪ fər mi ‘lie for me’	

Next, the child must undo phrase-level alternations by assigning a single representation to each word in the phrase-level input. At this stage, the learner does not yet attempt to analyse word-level collocations such as *writ-ing* and *rid-ing*; at the phrase level, these are treated in the same way as monomorphemic items like *mitre* and *powder*.

Let us first consider the alternation [hɪt] *hit* ~ [hɪr æn] *hit Ann*. If we assume minimum disparity between inputs and outputs (§4.4), there is only one possible phrase-level input representation for *hit*: viz. /hɪt/. Note that */hɪr/ and */hɪd/ would both incorrectly generate [hɪd]~[hɪr æn], as the phrase-level constraint hierarchy does not neutralize voice contrasts in word-final position. Crucially, by factoring out the identical portion of the alternants [hɪt]~[hɪr], the learner discovers a set of alternating elements [t]~[r]. And, given /hɪt æn/→[hɪr æn], she finds out that /t/ is a possible phrase-level input representation for [r] in the flapping environment.

Let us now turn to [hɪd] *hid* ~ [hɪr æn] *hid Ann*. Here, the set *J* of optimal phrase-level inputs for *hid* consists of two members: viz. /hɪd/ and */hɪr/ (§4.4). Since [d] and [r] are in complementary distribution on the surface, both representations generate the correct set of output alternants. In this case, however, both Hale’s heuristic (4a) and the heuristic for asymmetric paradigms (4b) favour input /hɪd/. Since the learner has no reason to retract this hypothesis, */hɪr/ is discarded. On this basis, the child discovers a new alternating set [d]~[r] derived from input /d/.

The child now knows that [r] in the Flapping environment is an archiphonemic string with two possible input correspondents: /t/ or /d/. By Archiphonemic Prudence (§4.5), therefore, non-alternating items such as *mitre*, *powder*, *writing*, and *riding* must be quarantined, and the assignment of phrase-level input representations to them is deferred. Assuming that the learner countenances the minimal departure from the identity map compatible with Archiphonemic Prudence, the choice of inputs will be as in (12):

(12)

<i>Quarantined item</i>	<i>Phrase-level input candidates</i>
[məɪrər] ‘mitre’	/məɪtər/, /məɪdər/
[paʊərər] ‘powder’	/paʊtər/, /paʊdər/
[rəɪrɪŋ] ‘writing’	/rəɪtɪŋ/, /rəɪdɪŋ/
[raɪrɪŋ] ‘riding’	/raɪtɪŋ/, /raɪdɪŋ/

¹⁰ Under M » F bias (see §4.3), it is preferable to impute violations of CLIPDIPH to a higher-ranked markedness constraint (i.e. CLEARDIPH), rather than to faithfulness (i.e. IDENT[low]).

5.3. Acquiring the word-level cophonology

Leaving aside the quarantined items in (12), the child can now proceed to the acquisition of the word-level cophonology. At this stage, the data set consists of the single whole words that remain in the non-quarantined phrase-level input: e.g. /hɪt/ *hit*, /hɪd/ *hid*, /rəɪt/ *write*, /raɪd/ *ride*, /aɪfʊl/ *eyeful*, etc. Crucially, there is no form in this data set where either [əɪ] or [ʌʊ] fails to be followed by a voiceless obstruent in the same foot. Recall that all items in which Raising overapplies word-internally, such as [məɪrər] *mitre* and [rəɪrɪŋ] *writing*, have been placed under quarantine. On the surface, Raising also overapplies in forms subject to Flapping across word boundaries: e.g. [rəɪr ʌp] *write up*. These forms, however, are involved in phrase-level alternations (e.g. [rəɪt] *write* ~ [rəɪr ʌp] *write up*) and consequently disappear in the processes of phrase-level input assignment. Remember that, at phrase level, [rəɪr ʌp] ← /rəɪt ʌp/ (see §5.2 again). Nonetheless, even if Raising no longer overapplies, there still remain instances of underapplication: e.g. [aɪfʊl] *eyeful*.

Let us now consider the outcome of pure phonotactic learning in this situation. Since the data include raised diphthongs, CLEARDIPH must be crucially dominated, either by CLIPDIPH or by IDENT[mid]. Note, however, that all violations of CLEARDIPH occur before voiceless obstruents in the same foot, for, as we have just seen, there is no overapplication of Raising in the non-quarantined data. Accordingly, a learner subject to M » F bias will respond to the datum rəɪt › raɪt by ranking CLIPDIPH above CLEARDIPH, whilst preserving the default ranking CLEARDIPH » IDENT[mid]. In contrast, the datum aɪfʊl › əɪfʊl cannot be imputed to a contextual markedness effect, and so triggers the ranking IDENT[low] » CLIPDIPH.¹¹

At this point, the quarantine on nonalternating items such as *mitre* and *writing* may be lifted, as the newly established word-level hierarchy forces a choice between the phrase-level input candidates allowed by Archiphonemic Prudence. Observe that *[məɪdər] and *[rəɪdɪŋ] are ill-formed word-level outputs because they show overapplication of Raising. These forms cannot therefore be derived from identical input under the word-level ranking IDENT[low] » CLIPDIPH » CLEARDIPH » IDENT[mid]. In consequence, the phrase-level input representations for *mitre* and *writing* must be /məɪtər/ and /rəɪtɪŋ/.

(13)

		IDENT[low]	CLIPDIPH	CLEARDIPH	IDENT[mid]
məɪdər	məɪdər			*!	
	mardər ➞				*
məɪtər ➞	məɪtər ➞			*	
	maɪtər		*!		*
rəɪdɪŋ	rəɪdɪŋ			*!	
	raɪdɪŋ ➞				*
rəɪtɪŋ ➞	rəɪtɪŋ ➞			*	
	raɪtɪŋ		*!		*

In contrast, it is not yet possible at this stage to lift the quarantine on *powder* and *riding*. In this case, the incorrect phrase-level inputs are */paʊtər/ and */raɪtɪŋ/, which contain

¹¹ The data are also compatible with less restrictive rankings such as IDENT[mid] » CLEARDIPH » CLIPDIPH. I assume, however, that, in line with the Subset Principle, the constraint ranking algorithm always selects the most restrictive hierarchy —although, admittedly, there are problems in trying to enforce the Subset Principle through an M » F bias (see Prince & Tesar 1999 for discussion).

unraised diphthongs followed by a voiceless obstruent in the same foot. However, since underapplication of Raising is still tolerated at the word level (cf. [aɪfʊl] *eyeful*), both these forms are possible word-level outputs. The choice of input for *powder* and *riding* must accordingly wait until the stem-level constraint hierarchy is known.

Nonetheless, the lifting of the quarantine on *mitre* and *writing* frees up more data for pure phonotactic learning at the word level. The new word-level output forms, e.g. [møitər] *mitre* and [røitɪŋ] *writing*, are counterexamples to Flapping and so enable the child to learn that Flapping does not apply at the word level (or higher in the grammar).

The child can now turn to input assignment. This is pretty straightforward at the word level, as the partial lifting of the quarantine has not revealed new alternations. Accordingly, the learner has no reason to deviate from the identity map: i.e. /hɪt/ → [hɪt], /hɪd/ → [hɪd], /møitər/ → [møitər], etc. In particular, word-level derivatives such as [aɪfʊl] *eyeful* and [røit-ɪŋ] *writing* do not create alternations with their respective base forms: cf. [aɪ] *eye* and [røit] *write*. The input representation of the stem will therefore be identical with its output realization: i.e. /aɪ-/ *eye* and /røit-/ *write*.

5.4. Acquiring the stem-level cophonology

By this time, the learner has taken a decisive step forward: in effect, when she removes word-level suffixes such as *-ful* and *-ship* from collocations such as [aɪfʊl] *eyeful* and [fraʊʃɪp] *Frauship* (see (10) above), she disposes of the last remaining instances of Raising misapplication. The input to the word level consists of (i) monomorphemic items such as /'møitər/ *mitre*, /røit/ *write*, /søit/ *cite*, /'søifn/ *syphon*, /aɪ/ *eye*, and (ii) stem-level collocations such as the irregular verbs /hɪt/ *hit* and /hɪd/ *hid* or the level-one derivatives /saɪ'fənɪk/ *syphonic* and /saɪ'teɪʃn/ *citation*. These forms, which provide the trigger for phonological acquisition at the stem level, obey Raising. In consequence, Raising becomes true of the stem-level output, and the appropriate constraint ranking can be installed in the stem-level hierarchy by pure phonotactic learning.

At last, the child can lift the quarantine on *powder* and *riding*. The newly acquired stem-level hierarchy successfully discards the incorrect phrase-level inputs */pəʊtər/ and */raɪtɪŋ/, where Raising underapplies. In consequence, /pəʊdər/ and /raɪdɪŋ/ are returned as the phrase-level input representations for *powder* and *riding*.

(14)

		CLIPDIPH	CLEARDIPH	IDENT[low]	IDENT[mid]
pəʊtər	pəʊtər	*!			
	pəʊtər →		*	*	
pəʊdər →	pəʊdər →				
	pəʊdər		*!	*	
raɪtɪŋ	raɪtɪŋ	*!			
	raɪtɪŋ →		*	*	
raɪdɪŋ →	raɪdɪŋ →				
	raɪdɪŋ		*!	*	

At the word level, the child can now sort out the paradigm [raɪd] *ride* ~ [raɪd-ɪŋ] *riding*. Since the paradigm proves to be non-alternating, the child adheres to the identity map and selects /raɪd-/ as the input representation of the stem. There then remains the

task of identifying the input to the stem level, but no special difficulty arises here.¹² For all intents and purposes, the acquisition of the counterbleeding interaction between Diphthong Raising and Flapping in Canadian English is now complete.

References

- Bermúdez-Otero, R. (1999). *Constraint interaction in language change [Opacity and globality in phonological change.]* PhD dissertation, University of Manchester / Universidad de Santiago de Compostela.
- Bermúdez-Otero, R. (2002) Cyclicity or sympathy? A case study. Handout of paper presented at the 10th Manchester Phonology Meeting.
- Bermúdez-Otero, R. (in preparation). *The life cycle of constraint rankings: studies in early English morphophonology.*
- Chomsky, N. (1964). Current issues in linguistic theory. In J. Fodor & J. Katz (eds), *The structure of language*. Englewood Cliffs, NJ: Prentice Hall. 50-118.
- de Lacy, P. (1998). Sympathetic stress. Ms., University of Massachusetts, Amherst. (ROA 294)
- Hale, K. (1973). Deep-surface canonical disparities in relation to analysis and change: an Australian case. In T. A. Sebeok (ed), *Current trends in linguistics*. Vol. 11: *Diachronic, areal, and typological linguistics*. The Hague: Mouton. 401-458.
- Hayes, B. (1999). Phonological acquisition in Optimality Theory: the early stages. Ms, UCLA (ROA 327). To appear in R. Kager, J. Pater & W. Zonneveld (eds), *Fixing priorities: constraints in phonological acquisition*. Cambridge: Cambridge University Press.
- Itô, J. & R. A. Mester (1997). Sympathy Theory and German truncations. In V. Miglio & B. Morén (eds), *Proceedings of the Hopkins Optimality Workshop/Maryland Mayfest 1997*. University of Maryland Working Papers in Linguistics **5**: 117-139.
- Inkelas, S. (1995). The consequences of optimization for underspecification. *Proceedings of the North East Linguistic Society* **25**: 287-302.
- Jensen, J. T. (2000). Against ambisyllabicity. *Phonology* **17**: 187-235.
- Joos, M. (1942). A phonological dilemma in Canadian English. *Language* **18**: 141-4.
- Kenstowicz, M. (1994). *Phonology in generative grammar*. Cambridge, MA: Blackwell.
- Kiparsky, P. (2000). Opacity and cyclicity. *The Linguistic Review* **17**: 351-365.
- McCarthy, J. J. (1999). Sympathy and phonological opacity. *Phonology* **16**: 331-399.
- McCarthy, J. J. (2003). Sympathy, cumulativity, and the Duke-of-York gambit. In C. Féry & R. v. d. Vijver (eds), *The syllable in Optimality Theory*. Cambridge: Cambridge University Press.
- Prince, A. & P. Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar*. Report no. RuCCS-TR-2. New Brunswick, NJ: Rutgers University Center for Cognitive Science. (Revised August 2002: ROA 537.)
- Prince, A. & B. Tesar (1999). Learning phonotactic distributions. Report no. RuCCS-TR-54. New Brunswick, NJ: Rutgers University Center for Cognitive Science. (ROA 353)
- Tesar, B. & P. Smolensky (2000). *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.
- Wells, J. C. (1982). *Accents of English*. Vol. 3: *Beyond the British Isles*. Cambridge: Cambridge University Press.

¹² This job involves undoing transparent alternations such as [səɪt] *cite* ~ [saɪ'teɪʃn] *citation* under Richness of the Base, just as in classical monostratal OT. See notes 1 and 7.

Pre-sonorant voicing in Slovak: the treatment of a derived environment effect in OT

Sylvia Blaho

Theoretical Linguistics Programme, Eötvös Loránd University
H-1068 Budapest, Benczúr u. 33, Hungary/
University of Leiden Centre for Linguistics
NL-2311BX Leiden, van Wijkplaats 4, The Netherlands
sylvia@nytud.hu

Szilárd Szentgyörgyi

Department of English and American Studies, University of Veszprém
H-8200 Veszprém, Egyetem u. 10, Hungary
szentsz@almos.vein.hu

Abstract. This paper presents voicing assimilation in Slovak, which seems to be sensitive to morphologically derived environments: word final obstruents become voiced before sonorant consonants and vowels if a strong boundary intervenes. We argue that an account combining traditional autosegmental representations with empty skeletal positions in an Optimality Theoretic (OT) grammar can correctly predict all the relevant forms in Slovak. We also demonstrate how this analysis is superior to a Stratal OT account (Kiparsky, 2000) or an account making use of a Derived Environment Constraint proposed by Polgárdi (1998).

1 Introduction

In this paper, we present voicing assimilation in Slovak, which seems to be sensitive to morphologically derived environments. Namely, word final voiceless obstruents become voiced if followed by a sonorant consonant or a vowel across a strong morpheme boundary. This peculiar behaviour is restricted to so-called analytical suffixation and it does not occur with synthetic suffixes, i.e. those that have no impact on phonology as argued by Kaye (1995).

Derived environment effects (DEs) have always been a challenge to non-derivational frameworks, especially Optimality Theory (OT), which, in its original form (Containment Theory as in Prince & Smolensky, 1993) is a purely output oriented theory of grammar. Although in a later version, the Correspondence Theory of Faithfulness (McCarthy & Prince, 1995), constraints referring to both inputs and outputs, so-called faithfulness constraints, were introduced, it is still impossible to refer to intermediate representations since they do not exist.

We propose an account of the Slovak voice assimilation data in the framework of Optimality Theory (OT, Prince & Smolensky 1993, McCarthy & Prince 1995), basing our analysis on the voicing typology suggested by Petrova et al. (2000) and incorporating the insights of a variety of Government Phonology (GP, Kaye et al. 1990) known as Strict CV Phonology (Lowenstamm 1996a,b, 1999; Scheer 2002, forth., Dienes & Szigetvári 1999, Szigetvári 1999). We argue that combining traditional autosegmental representations with empty skeletal positions in an OT-type of grammar, one can correctly predict all the relevant forms in Slovak. Also, we show how this analysis is superior to a Stratal OT account (Kiparsky, 2000) or an account in the form of OT proposed by Polgárdi (1998), who united GP representations with the OT machinery but claimed that DE effects can be treated by simply introducing a Derived Environment Constraint (DEC) requiring that there should be no change in non-derived environments.

The structure of the paper is as follows. In the second section, we present the relevant data and the generalisations drawn from them. In the third section, we describe the voicing typology proposed by Petrova et al (2000) and demonstrate how it has to be improved in order to be able to account for the new Slovak data and suggest an alternative analysis of the new data, using GP-type representations, in the fourth section. Finally, we discuss two proposed ways of dealing with DEs within OT, namely that of Polgárdi (1998) and Kiparsky (2000), and show why these proposals are to be dispreferred.

2 Data and generalisations¹

Slovak obstruent clusters always agree in voicing, their voiceless or voiced quality being determined by the rightmost obstruent of the cluster, as shown by the examples in (1a). In pre-pause positions only voiceless obstruents are allowed, i.e. Slovak displays final devoicing as in (1b). Within the phonological word (and also across weak morpheme boundaries), both voiceless and voiced obstruents occur before sonorant consonants and vowels, i.e. there is no assimilation to sonorants in this domain as illustrated in (1c). However, underlyingly voiceless obstruents appear as voiced on the surface when they precede a sonorant consonant or a vowel if a strong morpheme boundary separates the two (1d).

- | | | | |
|--------|-----------|-------------|--|
| (1) a. | pro[s]it' | pro[zb]a | 'ask' – 'request (n)' |
| | ža[b]a | ža[pk]a | 'frog' – id.dimin. |
| b. | pá[d]om | pá[t] | 'case Ins.Sg.' – id. Nom.Sg. |
| | br[zd]a | br[st] | 'break Nom.Sg.' – id. Gen.Pl. |
| c. | [st]rava | [zd]ravie | 'food' – 'health' |
| | [t]lak | [d]laň | 'pressure' – 'palm' |
| d. | voja[k]a | voja[g] ide | 'soldier Gen.Sg.' – '(the) soldier goes' |
| | p[s]a | pe[s] je | 'dog Gen.Sg.' – '(the) dog is' |

That is, besides the normal regressive voicing assimilation before obstruents in all environments, there is a further phenomenon to explain in Slovak: regressive voicing assimilation to sonorants (consonants and vowels) across strong morphological boundaries.

¹ The phenomenon discussed by Pauliny (1979) and Rubach (1996, 1997b), among others.

3 An OT typology of voicing: Petrova et al. (2000)

Petrova et al. (2000) propose the constraints in (2)-(5) to account for the voicing typology found cross-linguistically.

- (2) **Share** Obstruents in clusters must share laryngeal specifications.
- (3) **ID preson voice** A consonant in presonorant position must be faithful to the input specification for voice.
- (4) **ID voice** A consonant must be faithful to the input specification for voice.
- (5) ***voice** Voiced obstruents are prohibited.

In Slovak, these constraints must be ranked **Share, IDpreson voice** >> ***voice** >> **IDvoice** to account for the regular cases of voice assimilation and devoicing. If we assume this hierarchy, then we get both the devoicing and the regular regressive voice assimilation of Slovak.

(6)

a. bʲ/zd/	Share	ID.preson.voi	*voi	ID.voi
bʲ[zd]			*!	
☞ bʲ[st]				**
bʲ[zt]	*!		*	*
bʲ[sd]	*!		*	*

(7)

b. ža/bk/a	Share	ID.preson.voi	*voi	ID.voi
☞ ža[pk]a				*
ža[bk]a	*!		*	
ža[bg]a		*!	**	*
ža[pg]a	*(!)	*(!)	*	**

(8)

c. pro/sb/a	Share	ID.preson.voi	*voi	ID.voi
pro[sb]a	*!		*	
☞ pro[zb]a			**	*
pro[sp]a		*!		*
pro[zp]a	*(!)	*(!)	*	**

However, another constraint has to be at work that triggers presonorant voicing:

- (9) **Passive voice** Obstruents are voiced before sonorants.

This constraint has the desired effect of making an obstruent voiced before sonorants. However, it also has an undesirable effect: as a result, only voiced obstruent will surface in presonorant position. Another shortcoming of the same constraint is that it is unable to distinguish derived environments from non-derived ones. Unfortunately, the amended constraint hierarchy still fails to select the actual surface forms as optimal in the relevant cases as shown by an example in tableau (10):

(10)

/pes/ je	Share	Passive voice	IDpreson voice	*voice	IDvoice
[pes] je		*!*			
[pez] je		*!	*	*	*
⊗[bez] je			**	**	**

Since this failure to predict the optimal candidate is not only a characteristic of Petrova et al.'s analysis, but all analyses using such constraints, something else has to be suggested. A solution is presented in section 4 below.

4 Representations

4.1 Features

It has been suggested in the literature that the voicing of obstruents and sonorants is different – for instance, sonorants normally do not participate in voice assimilation and final devoicing – and that phonological representations should reflect this. In this paper we make use of representations used in Government Phonology, the so-called Element Theory (Harris 1990, 1994, Harris & Lindsey 1995, Szigetvári 1997, 1998), to make such a distinction. As this model utilises unary features, is more restricted than other representations using binary features.

Two features are relevant for our discussion: [voice], which functions roughly as [+voice], and [obst], which is more or less equivalent to the [-sonorant] feature of earlier models.² Szigetvári (1998) proposed that in obstruents, [voice] is a dependent of [obst] rather than being linked directly to a skeletal slot (11a). We propose that sonorant consonants and vowels also possess a feature [voice], linked directly to the skeletal slot (11b). Thus, the different behaviour of obstruents and sonorants in voicing phenomena is captured by the different structural positions of the feature [voice] in their representation.

(11)

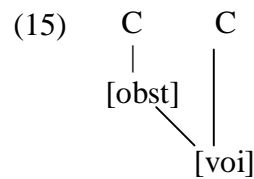
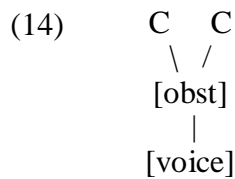
a. <i>obstruent voice:</i>	C	b. <i>sonorant voice:</i>	C	V
	[obst]		[voice]	[voice]
	[voice]			

On the basis of the above, we can now reformulate the constraint requiring that obstruent clusters agree in voicing and the one triggering presonorant voicing in the following way:

- (12) **Share** (reformulated) Obstruents adjacent on the skeleton must share their [obstruent] feature (and everything linked to it).
- (13) **Passive voice** (reformulated) The feature [voice] immediately dominated by a skeletal slot spreads onto a preceding [obstruent].

² In the GP literature, melodic primes are called **elements**. Instead of [obst] and [voice], the relevant primes are designated by **h** and **L**, respectively. We use a different notation in this paper because it might be more familiar to the non-GP audience.

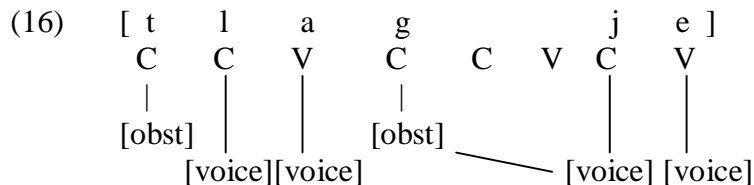
Accordingly, an obstruent cluster satisfying the constraint in (12) will look like (14), while a string of an obstruent plus a sonorant satisfying (13) will have the structure in (15).



4.2 The skeleton

As for the representation of boundaries, we consider boundaries as phonological entities along the lines of Strict CV Phonology (Lowenstamm 1996b, 1999, Scheer 1998, 2001, forth., Szigetvári 1999). The beginning of a word in this theory is represented by empty skeletal positions, a CV unit. This move is independently motivated by a range of phenomena such as cliticisation, initial consonant cluster phonotactics and liaison. As space restrictions do not permit us to present these arguments, the reader is referred to the above-mentioned references for details.

By representing a boundary with some skeletal positions, we are able to distinguish between voiceless obstruents followed by a sonorant within the same word and those that are followed by a sonorant across a strong boundary. An example of both configurations is shown in (16): the word initial [t] and the word-final [g] are both followed by a sonorant on the surface; however, while [t] and [l] are adjacent on the skeleton as well, [g] and [j] are separated by some empty positions.



While the constraints in (12) and (13) refer to features, positional faithfulness constraints such as **IDpreson voice** are evaluated by looking at the skeletal tier. Thus, in diagram (16), the surface obstruent [g] corresponding to an underlying [k] does not violate the constraint **IDpreson voice** as it is not in presonorant position since it is followed by an empty C position. The tableaux below show how our hierarchy with the modified constraints and representations is able to face the challenge of presonorant voicing.

(17)	tla/k/ je k C V j [o] [v]	ID preson Voice	Passive voice	*voice	IDvoice
	tla[k] je k C V j [o] [v]		*!		
	tla[g] je g C V j [o] [v]			*	*
(18)	tla/k/om k o [o] [v]	IDpreson Voice	Passive voice	*voice	ID voice
	tla[k]om k o [o] [v]		*		
	tla[g]om g o [o] [v]	*!		*	*

As tableau (17) shows, presonorant voicing is compulsory across an intervening strong boundary or else the output form violates **Passive voice** as the first candidate in (17). Within the phonological word, however, presonorant voicing is blocked by the faithfulness constraint, **IDpreson voice**. Note once again that this constraint is rendered inactive in (17) by the intervening CV skeletal slots between the stem final [g] and the word initial [j].

This way it is possible to explain why it is only across strong morpheme boundaries that underlyingly voiceless obstruents surface as voiced in presonorant position: it is because strong morpheme boundaries are always followed by empty skeletal positions, which protects the final consonant of the previous word from the effect of **IDpreson voice** and making it possible for it to become voiced under the effect of **Passive voice**.

5 Other solutions: the Derived Environment Constraint and Stratal OT

In this section, we review two other proposals for handling derives environment effects in OT: the Derived Environment Constraint of Polgárdi (1998) and Kiparsky's (2000) Stratal OT.

5.1 The Derived Environment Constraint

Following Kiparsky's (1973:9) Revised Alternation Condition, Polgárdi proposes the following constraint to prohibit neutralisation in non-derived environments, which we here take to refer to morphologically derived environments only:

- (19) **Derived Environment Constraint** No changes in non-derived environments.

According to Polgárdi (1998), this constraint has to be ranked between the one applying across the board and the other one, which only applies in derived environments as shown in the tableaux below.

(20)

/tlakom/	Share	DEC	Spread voi	IDpreson voi	*voice	IDvoi
[tlakom]			**			
[tlagom]		*!	*	*	*	*
[dlakom]		*!	*	*	*	*
[tlagom]		*!*		**	**	**

(21)

/pes/ je	Share	DEC	Spread voi	IDpreson voi	*voice	IDvoi
[pes] je			**!			
[pez] je			*	*	*	*
[bes] je		*!	*	*	**	*
[bez] je		*!		**	**	**

As it can be seen, if DEs are understood as referring to morphologically derived environments exclusively, then in (21), the second candidate will be selected as optimal as the last two candidates violate DEC while the first one violates Passive voice twice. We have to note, though, that there are two problems with this solution: on the one hand, it only considers morphological concatenation as derived environments and, on the other hand, the proposed constraint, DEC, cannot be classified either as a faithfulness, markedness or alignment constraint. As such, its status is highly problematic for a theory which only allows the above three basic types of constraints.

5.2 Stratal OT

Another suggestion to avoid DEs was proposed by Kiparsky (2000), called Stratal OT. It is basically a union of OT and Lexical Phonology (LP): Kiparsky argues that there are three strata in the phonology of a language, which correspond to the stem-level, word-level (together called lexical component) and phrase level phonology but each is parallel in the evaluation. That is, Kiparsky suggests that there are three strata in OT, each of which is a separate input-output device with parallel evaluation and the difference between the strata can only lie in the ranking of the constraints.

Let us see how this proposal helps us deal with the Slovak data:

(22)	/pes/	Share	IDpreson voice	*voice	IDvoice	Passive voice
	☞ [pes]					*
	[pez]			*!	*	*
	[bez]		*!	**	**	
(23)	/pes/	Share	IDpreson voice	*voice	IDvoice	Passive voice
	☞ [pes]					*
	[pez]			*!	*	*
	[bez]		*!	**	**	
(24)	/pes je/	Share	Passive voice	IDpreson voice	*voice	IDvoice
	[pes je]		*!*			
	[pez je]		*!	*	*	*
	⊗ [bez je]			**	**	**

As it can be seen, tableaux (22) and (23) show the stem-level and word-level phonology. The optimal output candidate in (22) is the input to (23), while the optimal output candidate in (23) is the input to (24). However, in (24) we have another word added after pes, and the constraints are also reranked. Passive voice has to dominate IDpreson voice to have any effect. Unfortunately, it is not able to distinguish the word initial /p/ and the word final /s/, and both will be voiced in the output, an unwelcome result. One way out would be to add Polgárdi's DEC, but that is problematic in itself as we have already noted above. Thus we can conclude that Stratal OT is no answer to our question, either.

6 Conclusion

In this paper, we analysed pre-sonorant voicing in Slovak, a phenomenon showing sensitivity to morphological boundaries. By making use of independently motivated devices such as feature geometry, unary features and empty skeletal positions, and combining them with the OT framework, we have successfully accounted for the problematic data.

References

- Dienes, Péter & Péter Szigetvári 1999. 'Repartitioning the skeleton: VC Phonology'. ms., Eötvös Loránd University, Budapest.
- Durand, Jacques & Francis Katamba (eds.) 1995. *Frontiers of Phonology: atoms, structures, derivations*. Longman, Harlow.
- Harris, John. 1990 'Segmental complexity and phonological government'. *Phonology* 7/2, 255-300.
- Harris, John 1994. *English Sound Structure*. Blackwell, Oxford & Cambridge, MA.
- Harris, John & Geoff Lindsey 1995. 'The elements of phonological representation'. In: Jacques Durand & Francis Katamba (eds.), 34-79.
- Kaye, Jonathan 1995. 'Derivations and interfaces'. In: Jacques Durand & Francis Katamba (eds.), 289-332.
- Kaye, Jonathan, Jean Lowenstamm and Jean-Roger Vergnaud 1990. 'Constituent structure and government in phonology'. *Phonology* 7/2 193-232.
- Kiparsky, Paul 1973. *Abstractness, Opacity and Global Rules*. Indiana University Linguistics Club.
- Kiparsky, Paul 1982a. 'Lexical Morphology and Phonology. In: I-S.Yang (ed.). *Linguistics in the Morning Calm*. Hanshin, Seoul, 3-91.
- Kiparsky, Paul 1982b. 'From Cyclic to Lexical Phonology'. In: Van der Hulst & Smith (eds.). *The Structure of Phonological Representations*, Part I. Foris, Dordrecht.
- Kiparsky, Paul 2000. 'Opacity and Cyclicity'. *The Linguistic Review* 17: 351-367.
- Lowenstamm, Jean 1996. 'CV as the only syllable type'. In Jacques Durand & Bernard Laks (eds.). *Current Trends in Phonology: Models and Methods. European Studies Research Institute, University of Salford Publications*, 419-442.
- Lowenstamm, Jean 1996. 'The phonological government of affixes. A theory of cliticization'. Paper presented at the 8th Vienna Phonology Meeting.
- Lowenstamm, Jean 1999. 'The beginning of the word'. In: John Rennison & Klaus Kühnhammer (eds.). *Syllables?!*, 153-166. Holland Academic Graphics, The Hague.
- McCarthy, John & Alan Prince 1995. 'Prosodic Morphology I. Constraint interaction and satisfaction'. ms., University of Massachusetts at Amherst and Rutgers University.
- Mohanan, Karuvannur Puthanveetil 1982. *Lexical Phonology*. Doctoral dissertation, Indiana University Linguistics Club.
- Pauliny, Eugén 1979. *Fonológia slovenského jazyka* [The phonology of Slovak]. Slovenské pedagogické nakladateľstvo, Bratislava.
- Petrova, Olga, Rosemary Plapp, Catherine O'Ringin & Szilárd Szentgyörgyi 2000. 'Constraints on voice: an OT typology'. Paper presented at the 2000 Meeting of the Linguistic Society of America.
- Polgárdi, Krisztina 1998. *Vowel Harmony. An account in terms of government and optimality*. PhD dissertation. Holland Academic Graphics, The Hague.
- Prince, Allan & Paul Smolensky 1993. 'Optimality Theory: Constraint Interaction in Generative Grammar'. ms. Rutgers university & University of Colorado at Boulder.
- Rubach, Jerzy 1996. 'Nonsyllabic analysis of voice assimilation in Polish'. *Linguistic Inquiry* 27, 69-110.
- Rubach, Jerzy 1997. 'Polish voice assimilation in Optimality Theory'. *Rivista di linguistica* 9/2, 291-342.
- Scheer, Tobias 1998. 'A Theory of Consonantal Interaction'. *Folia Linguistica* XXXII/3-4. 201-237.

- Scheer, Tobias 2001. 'The representation of morphological information in phonology'. Handout of course given at the 8th Eastern Generative Grammar Summer School in Niš, Yugoslavia.
<http://www.unice.fr/dsl/nis01/cvcv.htm>
- Scheer, Tobias 2002. 'How yers made Lightner, Gussman, Rubach, Spencer & Co invent CVCV'. ms., University of Nice.
<http://www.unice.fr/dsl/papers.htm>
- Scheer, Tobias (forth.). 'CVCV: a syntagmatic theory of phonology. On Locality, Morphology and Phonology in Phonology'. ms., University of Nice.
- Szigetvári, Péter 1997. 'Miért nem zöngésedik a [h]?' [Why [h] does not get voiced]. Paper presented at the A Mai Magyar Nyelv Leírásának Újabb Módszerei III. conference, April 16th-17th, Szeged, Hungary.
- Szigetvári, Péter 1998. 'Why [h] is not voiced'. In Eugeniusz Cyran (ed.) *Structure and interpretation. Studies in Phonology. PASE Studies & Monographs 4*. Lublin: Wydawnictwo Folium. 287-301.
- Szigetvári, Péter 1999. 'VC Phonology: a theory of consonantal interaction and phonotactics'. PhD dissertation, Eötvös Loránd Tudományegyetem/ MTA, Budapest.
<http://budling.nytud.hu/~szigetva/papers.html>

Overt forms and the control of comprehension

Paul Boersma

Institute of Phonetic Sciences, University of Amsterdam

Abstract. This paper shows that the commonly held serial view of the incorporation of overt forms in the grammar (e.g. Hayes 1996 for phonology, and Legendre, Smolensky & Wilson 1998 for syntax) is inconsistent with the even more commonly held view that if two distinct underlying forms are pronounced identically, at least one of them must violate faithfulness. By contrast, *perceptual control grammars* (Boersma 1998 for phonology, and Jäger 2002 for syntax) turn out to be consistent with this view of faithfulness.

1. Introduction

Optimality Theory claims to have replaced serial derivation with parallel evaluation. But when considering the inclusion of phonetic detail into the theory, most researchers revert to a serial view. For instance, Hayes (1996) admits: “Following Pierrehumbert (1980) and Keating (1985), I assume that there is also a phonetic component in the grammar, which computes physical outcomes from surface phonological representations. It, too, I think, is Optimality-theoretic [...]”. This testimony can be abbreviated as in (1), in which the arrows denote language-specific mappings, which can presumably be modelled as Optimality-Theoretic grammars (I will use the subscripts *u*, *s*, and *a* for underlying, surface, and articulatory forms, respectively).

(1) *The serial view of production in phonology*

$[underlying\ form]_u \rightarrow [surface\ form]_s \rightarrow [articulatory\ form]_a$

This is the prevailing view among phonologists who think that phonetic implementation should be modelled in the grammar at all. Syntacticians are a bit more than phonologists inclined to work with three representations, and a serial view of the grammar, as in (2), tends to be implicit in GB-style OT syntax (e.g. Legendre, Smolensky & Wilson 1998).

(2) *The serial view of production in syntax*

$[target\ logical\ form]_T \rightarrow [logical\ form]_L \rightarrow [phonetic\ form]_P$

In this paper, I will show that the serial view contradicts the very reason why OT-ists work with faithfulness constraints, which is summarized in (3).

(3) *The legitimacy of faithfulness*

If two different underlying forms are pronounced identically, at least one of their surface forms must violate a faithfulness constraint.

This axiom expresses the intuition that the way to formalize neutralization in OT is by punishing it with a faithfulness violation. I will assume the correctness of this assumption, because without it, faithfulness constraints would lose their indirect functional grounding.

If our interpretation of faithfulness is correct but incompatible with the serial view of the production grammar, it is the serial view that will have to go. I will replace it with (4).

(4) *The perceptual control view of the production grammar*

phonology: $[underlying]_u \rightarrow ([articulatory]_a \Rightarrow [auditory]_o \rightarrow [surface]_s)$

syntax: $[target]_T \rightarrow ([phonetic]_p \rightarrow [logical]_L)$

This perceptual control view reverts the order of all forms except the underlying form. The single arrows on the right stand for the reconstruction that the *listener* will be able to carry out on the message, and faithfulness constraints will be interpreted as evaluating (the speaker's view of) the extent to which the listener can reconstruct the message intended by the speaker. These recovery processes are language-specific and will therefore be modelled with Optimality-Theoretic grammars; the double arrow represents a language-independent process that therefore does not have to be modelled as a grammar.

Sections 2 to 5 will show how exactly the serial view goes wrong. Sections 6 to 8 will show that the control view does meet the legitimacy of faithfulness, and that it is the most natural view of OT production grammars that involve more than two representations.

2. Two representations, non-serial: McCarthy & Prince (1995)

Those versions of OT that work with only two representations have no fear of needing serial derivation. In Correspondence Theory (McCarthy & Prince 1995), the two representations are called *input* and *output*, but once one works with more than two representations, or studies both production and comprehension, such process-dependent labels are not sufficient, so I will instead use the more explicit traditional terms *underlying form* (UF) and *surface form* (SF). Tableau (5) shows how this version of OT models production.

(5) *McCarthy & Prince's formalization of production*

$[underlying]_u$	STRUCT _s	FAITH _{us}
$[surface_1]_s$		
$[surface_2]_s$		
$[surface_3]_s$		

Like the representations, the constraints are labelled with *u* and *s* in order to make explicit what representations they evaluate. Thus, the structural constraints, abbreviated here as STRUCT_s, evaluate aspects of the surface candidates only, while the faithfulness constraints, abbreviated here as FAITH_{us}, evaluate aspects of the similarity between the underlying form and the surface candidates (the order of STRUCT_s and FAITH_{us} in this schematic tableau has no relation to their relative ranking). An analogous tableau can be drawn for syntactic production with two representations (Legendre, Smolensky & Wilson 1998), in which the input is a *target form* (TF) and the output a *logical form* (LF). Such a tableau maps a $[target]_T$ to one of a number of candidates $[logical_i]_L$ via an evaluation of structural constraints at LF (STRUCT_L) and faithfulness constraints between TF and LF (FAITH_{TL}).

The two two-representation grammar models of production are summarized in (6).

(6) *Production models with two representations*

phonology: $[underlying]_u \rightarrow [surface]_s$

syntax: $[target]_T \rightarrow [logical]_L$

While I will need to modify the number of representations later on, I will assume that the faithfulness relation is defined correctly here. What is more, when introducing a third and

fourth representation I will continue to assume that SF is *defined* as the form whose similarity to UF is evaluated by faithfulness constraints. This definition allows us to derive from (3) an important intermediate result, formulated in (7).

(7) *The locus of neutralization*

If two different underlying forms are pronounced identically, this neutralization must occur somewhere in the mapping from underlying form to surface form.

We can see that this must be true by arguing that if the neutralization took place outside the path by which UF is mapped to SF, faithfulness constraints would not be able to evaluate it, hence (3) would be violated. There is, however, a small catch to this reasoning, as will become clear in the following section.

3. Three representations, non-serial: Tesar & Smolensky (2000)

The need for a third representation in phonology stems from the fact that language-learning children do not hear fully structured surface forms in their environment. Instead, they hear unstructured *overt forms*. For instance, when confronted with a sequence of three syllables, the second of which is stressed, they initially hear the overt form $[\sigma \acute{\sigma} \sigma]_o$ and have to *learn to construct* one of the surface forms $[(\sigma \acute{\sigma}) \sigma]_s$ or $[\sigma (\acute{\sigma} \sigma)]_s$, depending on whether their ambient language has iambic or trochaic feet. For this reason, Tesar & Smolensky (2000) propose a grammar model with three forms and two processes. Both mappings in (8) are language-specific, and they are handled by a single Optimality-Theoretic grammar.

(8) *Tesar & Smolensky's grammar model*

production: $[\text{underlying form}]_u \rightarrow [\text{full structural description}]_s$
 interpretation: $[\text{overt form}]_o \rightarrow [\text{full structural description}]_s$

The non-seriality of this grammar model relies heavily on *containment*, i.e., both the overt form and the underlying form are contained in the full structural description, see (9).

(9) *Non-serial grammar model with containment*

production: $[\text{underlying}]_u \rightarrow [\text{full description}]_s \Rightarrow [\text{overt}]_o$
 e.g. $[\sigma \sigma \sigma]_u \rightarrow [(\sigma \acute{\sigma}) \sigma]_s \Rightarrow [\sigma \acute{\sigma} \sigma]_o$ and $[\text{ta:g}+\emptyset]_u \rightarrow [\text{ta:g}_{\langle \text{voi} \rangle}+\emptyset]_s \Rightarrow [\text{ta:k}]_o$
 comprehension: $[\text{overt}]_o \rightarrow [\text{full description}]_s \Rightarrow [\text{underlying}]_u$
 e.g. $[\sigma \acute{\sigma} \sigma]_o \rightarrow [(\sigma \acute{\sigma}) \sigma]_s \Rightarrow [\sigma \sigma \sigma]_u$ and $[\text{ta:k}]_o \rightarrow [\text{ta:g}_{\langle \text{voi} \rangle}+\emptyset]_s \Rightarrow [\text{ta:g}+\emptyset]_u$

The second example in (9) is the nominative singular of the German word $[\text{ta:g}]_u$ ‘day’. The phonological part of the case ending is the null morpheme $[\emptyset]_u$. The word is pronounced with final devoicing and with aspiration of the initial voiceless plosive, i.e. as $[\text{t}^h\text{a:k}]_o$ (for the difference between this overt form and the one given by Tesar & Smolensky, i.e. $[\text{ta:k}]_o$, see below). The two double arrows in (9) are simple mechanical mappings. First, the mapping from the surface form to the overt form is mechanical, as summarized in (10).

(10) *Extracting the overt form from the full structural description*

- a. Delete hidden material such as parentheses, morphological boundaries, and null morphemes: $[(\]_s \Rightarrow [\]_o, [)]_s \Rightarrow [\]_o, [+]_s \Rightarrow [\]_o, [\emptyset]_s \Rightarrow [\]_o$
- b. Interpret the insertion and deletion marks: $[g_{\langle \text{voi} \rangle}]_s \Rightarrow [k]_o$

The mapping from the surface form to the underlying form is equally mechanical, as summarized in (11).¹

¹ Tesar & Smolensky (2000: 79) actually give $[\text{ta:g}_{\langle \text{voi} \rangle}]_s$ rather than $[\text{ta:g}_{\langle \text{voi} \rangle}+\emptyset]_s$ for the full structural description, thereby violating containment.

(11) *Extracting the underlying form from the full structural description*

- a. Delete metrical parentheses and stress marks: $[\langle \rangle_s \Rightarrow \langle \rangle_u, \langle \rangle_s \Rightarrow \langle \rangle_u, [\acute{\sigma}]_s \Rightarrow [\sigma]_u$
- b. Delete the insertion and deletion marks: $[g_{\langle \text{voi} \rangle}]_s \Rightarrow [g]_u$

We can now see that (7) does not necessarily follow from (3). Consider the German underlying forms $[\text{ra:d}+\emptyset]_u$ ‘wheel-NOMSG’ and $[\text{ra:t}+\emptyset]_u$ ‘advice-NOMSG’, both of which are pronounced $[\text{ʁa:t}]_o$, i.e., the final obstruent voicing contrast is neutralized. The full structural descriptions are $[\text{ra:d}_{\langle \text{voi} \rangle}+\emptyset]_s$ and $[\text{ra:t}+\emptyset]_s$, respectively. In the style of the containment faithfulness constraints of Prince & Smolensky (1993), the first of these forms violates PARSE (voi), while the second violates no faithfulness constraints at all. This means that metarule (3) is satisfied. But metarule (7) is not: the two surface forms have different structures, so the neutralization must take place in the mapping from SF to OF, i.e. in the steps $[d_{\langle \text{voi} \rangle}]_s \Rightarrow [t]_o$ and $[t]_s \Rightarrow [t]_o$. In other words, the neutralization takes place *after* it has been evaluated by the faithfulness constraints. To prevent this counter-intuitive situation, one would have to introduce the separate metarule in (12).

(12) *The anti-diacritical metarule*

Processes are evaluated where they are implemented.

If this metarule is assumed, (7) does follow from (3). We must note that (12) is incompatible with the containment view of the surface form: in order to prevent neutralization from being implemented after its evaluation, surface forms should contain $[t]_s$ rather than $[d_{\langle \text{voi} \rangle}]_s$, and if morpheme boundaries and null morphemes are subject to faithfulness as well, surface forms should not contain any instances of $[+]_s$ or $[\emptyset]_s$ either. This idea was implemented in later developments of Optimality Theory, as described in the next section.

4. Three representations, serial: Correspondence Theory with overt forms

Correspondence Theory (McCarthy & Prince 1995) is the OT dialect that assumes the anti-diacritical metarule (12). The surface form no longer contains insertion or deletion symbols or morphological information. This does not mean that all hidden material is erased: metrical structure is traditionally kept, since it is often hard to imagine how stress assignment can be handled without reference to hidden foot structure. The grammar model now turns into (13).

(13) *Serial grammar model with correspondence*

- production: $[\text{underlying}]_u \rightarrow [\text{surface}]_s \rightarrow [\text{overt}]_o$
 e.g. $[\sigma \sigma \sigma]_u \rightarrow [(\sigma \acute{\sigma}) \sigma]_s \rightarrow [\sigma \acute{\sigma} \sigma]_o$ and $[\text{ta:g}+\emptyset]_u \rightarrow [\text{ta:k}]_s \rightarrow [\text{t}^h\text{a:k}]_o$
 comprehension: $[\text{overt}]_o \rightarrow [\text{surface}]_s \rightarrow [\text{underlying}]_u$
 e.g. $[\sigma \acute{\sigma} \sigma]_o \rightarrow [(\sigma \acute{\sigma}) \sigma]_s \rightarrow [\sigma \sigma \sigma]_u$ and $[\text{t}^h\text{a:k}]_o \rightarrow [\text{ta:k}]_s \rightarrow [\text{ta:g}+\emptyset]_u$

The overt form is represented here with aspiration, unlike in (9), since German-learning children cannot a priori decide whether German aspiration is allophonic or not; for their part, they may well be learning a language with an underlying triple contrast between voiced, voiceless, and aspirated plosives, in which case aspiration is crucial. This general criticism of the view in (9) renders the SF-to-OF mappings in (9) and (13) non-mechanical. The change in SF between (9) and (13) renders the SF-to-UF mappings in comprehension non-mechanical as well, since the surface form $[\text{ra:t}]_s$ should now be mapped to either $[\text{ra:d}+\emptyset]_u$ or $[\text{ra:t}+\emptyset]_u$, probably depending on the semantic and pragmatic context. Both the SF→OF and SF→UF mappings have now become non-trivial, so that both the production and the comprehension process must be regarded as consisting of two serially ordered subprocesses. For production, we can identify these processes as *phonology* and *phonetic implementation*, and for comprehension, they are *perception* and *recognition*.

If the order of the subprocesses in the production model in (13) is correct, (7) reduces to the very simple statement in (14).

(14) *The non-neutralization of phonetic implementation*

The mapping from surface to overt form does not neutralize.

Despite its simplicity, (14) turns out to be extremely difficult to enforce, because it conflicts with the requirement that faithfulness constraints should be able to evaluate the UF-SF similarity. To see this, we consider two extreme interpretations of what a surface form is.

The first possible interpretation for SF is that it is a rather abstract form consisting of the same kind of discrete elements as UF. Under such an interpretation, the SF in (13) is [ta:k]_s, and its similarity to UF is easy to evaluate: it violates IDENT_{us} (voi) because the underlying segment [g]_u is voiced and its corresponding surface segment [k]_s is not; the remaining parts of the underlying form, [t]_u and [a:]_u, surface perfectly. While faithfulness constraints work well under this interpretation of SF, the non-neutralization of the SF→OF mapping cannot be guaranteed: who can tell whether the aspiration of [t]_s causes neutralization or not? Presumably it does not in German, but consider a couple of allophonic rules in Sanskrit and Japanese. In Sanskrit, an underlying [s]_u surfaces as [h]_o utterance-finally. Since the voiceless *[h]_u is not a possible lexical segment in Sanskrit, this must be regarded as an allophonic rule, hence [s]_u → [s]_s → [h]_o. However, an underlying [r]_u surfaces as [h]_o utterance-finally as well, hence [r]_u → [r]_s → [h]_o. But this is impossible, because it would mean that both [s]_s and [r]_s neutralize into [h]_o during phonetic implementation. A similar case occurs in Japanese, where [z]_u turns into the allophonic affricate [dʒ]_o before [i]_s, hence [z+i]_u → [zi]_s → [dʒi]_o, but [d]_u undergoes the same change, hence [d+i]_u → [di]_s → [dʒi]_o, again showing neutralization in phonetic implementation. These two cases of neutralization would leave faithfulness constraints powerless: despite the neutralization of [s]_u and [r]_u in Sanskrit, or [z+i]_u and [d+i]_u in Japanese, no faithfulness constraints are violated, since the surface forms are identical to the underlying forms. To be true, this situation could be patched up: unnatural derivations like [r]_u → [s]_s → [h]_o and [d+i]_u → [zi]_s → [dʒi]_o would do the trick of violating faithfulness by moving the neutralization to the UF→SF mapping, but the complication of the additional two unnatural changes (r→s and d→z) is something most phonologists nowadays would prefer to avoid. Precisely this type of complications was the reason for Halle (1959) to propose that an intermediate form (SF) does not exist. This is the standpoint taken by Chomsky & Halle (1968), according to whom the grammar maps UF to OF via a potentially large number of intermediate representations, none of which has any special status. Chomsky & Halle can be regarded as taking the opposite viewpoint from the abstract-SF viewpoint discussed above: for them, SF is the same as OF, and it is maximally rich. Such a situation does work fine for the requirement of non-neutralization of phonetic implementation, but a phonetically rich SF cannot be used by faithfulness constraints. There is no simple way in which the similarity of a discrete UF with a phonetically detailed SF could be evaluated: does [t^ha:k]_s violate DEP (aspiration) or not? If faithfulness constraints are to have any meaning at all, the underlying and surface forms should be *commensurable*, i.e., they should consist of the same kind of elements.

It seems that we have too many requirements for SF. For commensurability with UF, SF should be maximally abstract, but in order to make sure that the faithfulness constraints capture all cases of neutralization, SF should be maximally rich. This is probably why a worked-out serial theory of the production grammar, as summarized in (13), has never been proposed. While the issues tackled in the Correspondence Theory literature can often bear agnosticism with respect to the problems with serialism, phonetically-oriented dialects of OT cannot get by without facing these problems, as I will discuss in the following section.

5. Phonetic detail, serial

Phonetically inspired theories of phonology have to make a principled distinction between two overt forms: an articulatory form and an auditory form (Boersma 1989, Flemming 1995, Steriade 1995, Hayes 1996, Kirchner 1998). It is natural to assume that the speaker will produce an articulatory form and that the listener will start from an auditory form. The serial grammar model of (13) will turn into (15), although none of the works cited makes this proposal more explicit than the footnote from Hayes (1996) that I quoted in the Introduction above. I will label articulatory forms with *a*, and continue to label auditory forms with *o*.

(15) *Serial grammar model with phonetic detail*

production: $[underlying]_u \rightarrow [surface]_s \rightarrow [articulatory]_a$
 e.g. $[\sigma \sigma \sigma]_u \rightarrow [(\sigma \acute{\sigma}) \sigma]_s \rightarrow [\sigma \acute{\sigma} \sigma]_a$ and $[ta:g+\emptyset]_u \rightarrow [ta:k]_s \rightarrow [t^h\grave{a}:k]_a$
 comprehension: $[auditory]_o \rightarrow [surface]_s \rightarrow [underlying]_u$
 e.g. $[\sigma \acute{\sigma} \sigma]_o \rightarrow [(\sigma \acute{\sigma}) \sigma]_s \rightarrow [\sigma \sigma \sigma]_u$ and $[t^h\grave{a}:k]_o \rightarrow [ta:k]_s \rightarrow [ta:g+\emptyset]_u$

In (15), I have regarded the commensurability requirement as more important than the non-neutralization requirement. After all, one could still require that the phonetic implementation subprocess is non-neutralizing, perhaps by a smart technical invention. But that is not how I will handle the problem, because one can observe here a *conspiracy*: the technical details of a formalization of phonetic implementation would have to conspire in such a way that it does not map two distinct SFs to the same OF. As we learned from Prince & Smolensky (1993), whenever there seems to be a conspiracy there must be something wrong with the theory.

6. Phonetic detail, non-serial

I propose that the thing that is wrong with the theory in (15) is the serial $UF \rightarrow SF \rightarrow AF$ mapping, and more in particular the supposedly non-neutralizing $SF \rightarrow AF$ mapping. We can observe that there is nothing wrong with the *reverse* mapping, $OF \rightarrow SF$, which occurs in (15) as well. For instance, the $OF \rightarrow SF$ mapping is typically neutralizing, as can be expected from any mapping without conspiring requirements. Thus, the continuous detailed auditory form $[t^h\grave{a}:k]_o$ will be perceived as the segment sequence $[ta:k]_s$, but $[t^h\grave{a}:k]_o$ will also be perceived as $[ta:k]_s$, since German allows some variation in the place of the long low vowel. Some things nearby will be perceived differently: both $[d\grave{a}:k]_o$ and $[t\grave{a}:k]_o$ will be perceived as $[da:k]_s$ because German usually devoices its initial ‘voiced’ plosives, and both $[t^h\grave{a}ek]_o$ and $[t^h\grave{e}:k]_o$ will be perceived as the segment sequence $[tark]_s$ because German $[r]_u$ is vocalized as a lower mid central vowel when appearing in the coda of a syllable, often influencing the preceding vowel. From the literature, we know that OT grammars typically cause some cases of neutralization to occur. It is natural, therefore, to model the $OF \rightarrow SF$ mapping in OT (as a *perception grammar*, Boersma 1998), but it is unnatural to try to model $SF \rightarrow AF$ in OT.

If phonetic implementation cannot be modelled in OT, and it is still language-specific (as the examples show), the question remains whether it should be modelled at all. I propose that it should not. Instead, the reverse mapping, $OF \rightarrow SF$, which is needed in comprehension anyway, should take its place. We obtain the grammar model in (16).

(16) *Perceptual control view of phonological production*

production: $[underlying]_u \rightarrow ([articulatory]_a \Rightarrow [auditory]_o \rightarrow [surface]_s)$
 e.g. $[ta:g+\emptyset]_u \rightarrow ([t^h\grave{a}:k]_a \Rightarrow [t^h\grave{a}:k]_o \rightarrow [ta:k]_s)$
 comprehension: $[auditory]_o \rightarrow [surface]_s \rightarrow [underlying]_u$
 e.g. $[t^h\grave{a}:k]_o \rightarrow [ta:k]_s \rightarrow [ta:g+\emptyset]_u$

The second single arrow after ‘production’ is not phonetic implementation, but its reverse, namely perception. The idea is that the speaker chooses an articulation (AF) whose auditory

result (OF) will be perceived by the listener as a form (SF) that is as similar as possible to the speaker's intended message (UF), given the articulatory constraints. In other words, the objective of the speaker is to *control* the listener's perception, in the same sense in which Powers (1973) argued that *all* behaviour serves the control of perception. The double arrow in (16) is the mapping from articulatory form to auditory form; this is a language-independent mapping that involves physical (acoustical) and physiological transmissions.

The grammar model in (16) satisfies all three requirements (3), (7), and (12). If two different UFs are pronounced in the same way, i.e., if they have identical articulatory and auditory forms, the corresponding SFs will be identical as well; the direction of the arrows ensures this, since an OT grammar will always yield the same output for the same input as long as the ranking of the constraints does not change; hence, (7) is satisfied. Metarule (3) is then also satisfied, because a single SF cannot be identical to two different UFs at the same time. Metarule (12) has become irrelevant, since diacritics cannot pass from UF to AF, let alone to SF (though it is not impossible that the perception process *constructs* some default morphological information, e.g. that the SF in (16) is really [ta:k+Ø]_s).

The interpretation of what a faithfulness constraint is, has changed now: *faithfulness constraints evaluate (the speaker's view of) the extent to which the listener will be able to reconstruct the intended message without lexical access*. The interpretation of what phonetic implementation is, has also changed: *phonetic implementation does not exist as a module of the grammar*. Analogously to (16), (17) proposes a control grammar model for syntax.

(17) *The control view of syntactic production*

production: [target]_T → ([phonetic]_P → [logical]_L)
comprehension: [phonetic]_P → [logical]_L → [target]_T

7. The control view of the candidate generator

The parentheses around AF⇒OF→SF in (16) mean that the production grammar has to find the optimal triplet of AF-OF-SF combinations. In the same production grammar, constraints on articulatory effort evaluate the articulatory form (AF), structural constraints evaluate the surface form (SF), and faithfulness constraints evaluate the similarity of the surface form to the underlying form (UF). Instead of (5), tableaux will look like (18).


(18) *The control view of a production tableau*

[underlying] _u	ART _a	STRUCT _s	FAITH _{us}
[art ₁] _a ⇒ [aud ₁] _o → [surf ₁] _s			
[art ₂] _a ⇒ [aud ₂] _o → [surf ₂] _s			
[art ₃] _a ⇒ [aud ₃] _o → [surf ₃] _s			

The single arrow in each cell means that SF has to be computed from OF in a language-specific way, without reference to UF. This makes it impossible to have two candidates in which the auditory forms are identical but the surface forms are not.


Tableau (19) shows how the German neutralization example works in this model.

(19) *The control view of neutralization*

[ra:d+Ø] _u	NOFINAL VOICED OBSTRUENT _a	IDENT _{us} (voi)
[ɕa:d] _a ⇒ [ɕa:d] _o → [ra:d] _s	*!	
 [ɕa:t] _a ⇒ [ɕa:t] _o → [ra:t] _s		*

In such simple cases, the control view works similarly to Correspondence Theory. The Sanskrit case of multiple sources for the [h]_o allophone is more interesting. Consider the UF [maṭṭar]_u ‘mother’, which is pronounced as [maṭṭəh]_a. The question is to what extent the listener can reconstruct the underlying form from the auditory form [maṭṭəh]_o. Since all overt instances of [ə]_o derive from an underlying [a]_u (throughout Sanskrit phonology this vowel acts as the short counterpart to [aṃ]_u), the listener will have no problems in perceiving [ə]_o as [a]_s. The case is more difficult for [h]_o. Since the lexicon does not contain any instances of voiceless [h]_u, there is no point in perceiving [h]_o as [h]_s. On average, the listener will do better in reconstructing intended messages if she notes that the great majority of instances of [h]_o in Sanskrit derive from an underlying [s]_u (final [r]_u is far less common). The tableau in (20) shows how the listener will therefore perceive [maṭṭəh]_o as [maṭṭas]_s.


(20) *The perception of an overt voiceless glottal fricative in Sanskrit*

[maṭṭəh] _o	*[ə] _s	*[h] _s	[h] _o is not [k] _s	[ə] _o is not [i] _s	[h] _o is not [r] _s	[h] _o is not [s] _s	[ə] _o is not [a] _s
[maṭṭəh] _s	*!	*					
[maṭṭar] _s					*!		
 [maṭṭas] _s						*	*
[maṭṭis] _s				*!			
[maṭṭak] _s			*!				

We see that the perception process can be modelled in OT quite well. The constraints in (20) have been modelled in the style of Escudero & Boersma (2001). The constraints against perceiving [ə]_o as anything but [a]_s or against perceiving [h]_o as anything but [s]_s must be ranked high. In particular, it must be worse to perceive [h]_o as [r]_s than to perceive it as [s]_s. Escudero & Boersma show that such rankings automatically emerge during lexicon-driven acquisition as a result of different likelihoods, i.e., for the overt form [h]_o the candidate [s]_s is more likely to be ‘correct’ than the candidate [r]_s, since the learner is more likely to find [s]_u than [r]_u in her lexicon afterwards during recognition. Finally, the constraints *[ə]_s and *[h]_s must be ranked high, since such structures do not occur in the lexicon (alternatively, the candidate generator might not generate candidates with such structures in the first place, in which case we could do without these constraints).

We can now construct the production tableau for [maṭṭar]_u, as in (21). For brevity, the two overt forms (articulatory and auditory) have been collapsed into one, labelled *ao*.

(21) *The control view of neutralization into a distant allophone*

[maṭṭar] _u	NOFINALRHOTIC _a	IDENT _{us} (son)
[maṭṭər] _{ao} → [maṭṭar] _s	*!	
 [maṭṭəh] _{ao} → [maṭṭas] _s		*

Since it is optimal for the listener to map [maṭṭəh]_o to [maṭṭas]_s, there is no candidate like [maṭṭəh]_{ao} → [maṭṭar]_s. Thus, a given AF can never appear twice in the same tableau. In the formulation by Jäger (2002) for syntax, all candidates in production tableaux must be ‘hearer-optimal’. This is crucial in this case, since if we had been allowed to include the candidate [maṭṭəh]_{ao} → [maṭṭar]_s, it would have become the winning candidate since it violates none of the relevant constraints. In the same vein, two of the six candidates in


tableau 15 of Legendre, Smolensky & Wilson (1998) would not be generated in a control view of a syntactic production grammar, since their phonetic forms are identical to those of two hearer-optimal candidates (this *might* help solving one of the problems that they note...).

Interestingly, we see in (21) that the intermediate representation in the $[r]_u \rightarrow [s]_s \rightarrow [h]_o$ mapping discredited in §4 now reappears in the mapping $[r]_u \rightarrow ([h]_o \rightarrow [s]_s)$. In the present case, however, the occurrence of $[s]_s$ is not inspired by a metalinguistic need to prevent neutralization in phonetic implementation, but by the most sensible guess for Sanskrit listeners.

8. How control grammars incorporate phonetic detail


Since the control view of the production grammar does not allow a separate component for phonetic implementation, it remains to be shown how it is capable of expressing language-specific needs for certain phonetic details. As an example, tableau (22) shows how the aspiration in the initial plosive in the German $[ta:t + \emptyset]_s$ ‘deed-NOMSG’ comes about.

(22) *The control view of the implementation of phonetic detail*

$[ta:t + \emptyset]_u$	IDENT _{us} (voi / 96%)	IDENT _{us} (voi / 80%)	IDENT _{us} (voi / 20%)	*ASP _a	*LAX _a	IDENT _{us} (voi / 4%)
 $[t^h a:t]_{ao} \rightarrow 95\% [ta:t]_s, 5\% [da:t]_s$				*		*
$[t a:t]_{ao} \rightarrow 40\% [ta:t]_s, 60\% [da:t]_s$			*!			*
$[d a:t]_{ao} \rightarrow 10\% [ta:t]_s, 90\% [da:t]_s$		*!	*			*
$[d a:t]_{ao} \rightarrow 2\% [ta:t]_s, 98\% [da:t]_s$	*!	*	*		*	*

If constraints are ranked along a continuous scale, and some noise is added to the rankings at evaluation time (Boersma & Hayes 2001), the output of the perception grammar will vary from evaluation to evaluation. Hence, each of the four candidates has certain probabilities of being perceived as $[ta:t]_s$ and as $[da:t]_s$. For instance, the voiceless unaspirated articulation $[t a:t]_a$ is ambiguously perceived as $[ta:t]_s$ 40% of the time, as $[da:t]_s$ 60% of the time. I assume that the speaker knows these percentages (to compute them, she could run $[t a:t]_o$ through her perception grammar a number of times) and that the production grammar contains constraints that refer to them. For instance, $[t a:t]_{ao}$ violates IDENT_{us} (voi / 20%) because the probability that this candidate is perceived as the faithfulness-violating $[da:t]_s$ is more than 20%. Since it is worse to violate IDENT (voi) 80% of the time than it is to violate it only 20% of the time, the tableau exemplifies a fixed ranking by confusion probability. The tableau also contains a couple of articulatory constraints, which express the idea that it costs some effort to either aspirate a plosive, as in $[t^h]_a$, or to render it fully voiced, as in $[d]_a$.

(23) *The control view of the implementation of phonetic detail*

$[dax + \emptyset]_u$	IDENT _{us} (voi / 96%)	IDENT _{us} (voi / 80%)	IDENT _{us} (voi / 20%)	*ASP _a	*LAX _a	IDENT _{us} (voi / 4%)
$[t^h ax]_{ao} \rightarrow 95\% [tax]_s, 5\% [dax]_s$		*!	*	*		*
$[t ax]_{ao} \rightarrow 40\% [tax]_s, 60\% [dax]_s$			*!			*
 $[d ax]_{ao} \rightarrow 10\% [tax]_s, 90\% [dax]_s$						*
$[d ax]_{ao} \rightarrow 2\% [tax]_s, 98\% [dax]_s$					*!	

The same ranking explains the pronunciation of [d]_u as lenis voiceless, exemplified in tableau (23) for the underlying form [dax]_u ‘roof’. In this case, the candidate that would serve the listener best (namely [dax]_{ao}) fails to win, because the speaker does not bother to trade the articulatory gain of not performing the obstruent voicing gestures for an only slightly lower probability of confusion.

9. Conclusion

Unlike theories that propose a serial modularity of phonology and phonetic implementation, the perceptual control view of Optimality-Theoretic production grammars allows us to use faithfulness constraints for the purpose that they were designed for (including the evaluation of neutralization) and in the way they were defined by Correspondence Theory (namely as evaluating two commensurable discrete representations), while at the same time it allows us to explain the details of continuous phonetic implementation.

References

- Boersma, Paul (1989). Modelling the distribution of consonant inventories by taking a functionalist approach to sound change. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* **13**: 107–123.
- Boersma, Paul (1998). *Functional phonology*. PhD thesis, University of Amsterdam. The Hague: Holland Academic Graphics.
- Boersma, Paul, & Bruce Hayes (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* **32**: 45–86.
- Chomsky, Noam, & Morris Halle (1968). *The sound pattern of English*. New York: Harper and Row.
- Escudero, Paola, & Paul Boersma (2001). Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. *Rutgers Optimality Archive* **439**, <http://roa.rutgers.edu>. To appear in *Proceedings of the 25th Penn Linguistics Colloquium*.
- Flemming, Edward (1995). *Auditory representations in phonology*. PhD thesis, UCLA.
- Halle, Morris (1959). *The sound pattern of Russian*. The Hague: Mouton.
- Hayes, Bruce (1996). Phonetically-driven phonology: The role of Optimality Theory and inductive grounding. *Rutgers Optimality Archive* **158**, <http://roa.rutgers.edu>. Published in 1999 in Michael Darnell, Edith Moravcsik, Michael Noonan, Frederick Newmeyer, & Kathleen Wheatley (eds.) *Functionalism and formalism in linguistics*, Vol. I: *General papers*, 243–285. Amsterdam: John Benjamins.
- Jäger, Gerhard (2002). Learning constraint sub-hierarchies: The Bidirectional Gradual Learning Algorithm. *Rutgers Optimality Archive* **544**, <http://roa.rutgers.edu>. To appear in Henk Zeevat & Reinhard Blutner (eds.) *Optimality Theory and pragmatics*. Palgrave Macmillan.
- Kirchner, Robert (1998). *Lenition in phonetically-based Optimality Theory*. PhD thesis, UCLA.
- Legendre, Géraldine, Paul Smolensky, & Colin Wilson (1998). When is less more? Faithfulness and minimal links in wh-chains. In Pilar Barbosa, Danny Fox, Paul Hagstrom, Martha McGinnis, & David Pesetsky (eds.) *Is the best good enough?: optimality and competition in syntax*, 249–289. Cambridge, Mass.: MIT Press.
- McCarthy, John, and Alan Prince (1995). Faithfulness and reduplicative identity. In Jill Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*. *University of Massachusetts Occasional Papers* **18**, 249–384. Amherst, Mass.: Graduate Linguistic Student Association.
- Powers, William T. (1973). *Behavior: The control of perception*. Chicago: Aldine.
- Prince, Alan, & Paul Smolensky (1993). *Optimality Theory: Constraint interaction in generative grammar*. Technical Report TR-2, Rutgers University Center for Cognitive Science.
- Steriade, Donca (1995). *Positional neutralization*. Unfinished ms, Department of Linguistics, UCLA.
- Tesar, Bruce, & Paul Smolensky (2000). *Learnability in Optimality Theory*. Cambridge, Mass.: MIT Press.

Opacity and transparency related to lowering: Local Conjunction or Comparative Markedness^{*}

Kan Sasaki

Sapporo Gakuin University, Ebetsu

Abstract. A situation where opaque and transparent interactions related to lowering co-exist is found in the Mitsukaido dialect of Japanese. Local Conjunction can deal with this situation while Comparative Markedness cannot. The key difference lies in the treatment of derived structures from distinct processes.

1. Introduction

The aim of this paper is two-fold: to provide an Optimality Theoretic account for the opaque and transparent interactions among the phonological processes concerning lowering in the Mitsukaido dialect of Japanese (MD), and through the analysis, to show that Local Conjunction (Smolensky 1995) and Comparative Markedness (McCarthy 2002), two theoretical extensions of Optimality Theory (OT) both of which are regarded as useful devices for dealing with counterfeeding opacity, do not give equivalent results for a certain type of opaque phonological interaction. The analysis will suggest that Local Conjunction is an available extension for the problem while Comparative Markedness is not.

Constraints in OT are classified into two categories, namely markedness constraints and faithfulness constraints. Local Conjunction is an extension applicable to both markedness constraints and faithfulness constraints. The account for counterfeeding opacity in terms of local conjunction involves the use of conjoined faithfulness constraints. On the other hand, Comparative Markedness is an extension of markedness constraints. The theoretical extension with Local Conjunction and the one with Comparative Markedness look in opposite directions. The failure of Comparative Markedness to account for the MD interaction reveals a limit in the applicability of the markedness-based extension for a situation including counterfeeding opacity. This paper will also clarify the source of the inadequacy of the markedness-based extension to account for a situation where opacity and transparency related to the same process co-exist.

^{*} The data used in this paper is based on my field research and on the previous literature (Miyajima 1961). I am grateful to Mr. Nisaku Otaki, for answering questions patiently. Thanks also go to Daniela Caluianu and an anonymous reviewer for valuable comments. This research is supported by the Sapporo Gakuin University Research Support Grant (SGUS0220100603). All errors and shortcomings are my own.

2. Interactions of three phonological processes in the Mitsukaido dialect

The MD, spoken in the southwestern part of the Ibaraki prefecture, has a number of phonological and morpho-syntactic properties distinguishing it from Standard Japanese (SJ).¹ This paper will concentrate on the interaction among three phonological processes, namely coalescence, lowering, and /w/-deletion. This section begins with a brief description of vowels and glides, which are the targets of the processes discussed in this paper.

The dialect has 5 vowels /i, e, a, o, u/² and two glides /j/ and /w/ distinguished in terms of backness.³ The phonotactic distribution of the glide /j/ in MD is restricted compared to SJ. It can be followed only by [-high, -back] vowels.

(1)	SJ:	*ji	ju	MD:	*ji	*ju
		*je	jo		*je	jo
		ja			ja	

As well as in SJ, the only vowel which can follow the glide /w/ is /a/.

(2) *wi, *we, wa, *wo, *wu

When any other vowel follows, /w/ is deleted. This phenomenon can be observed in the verbal paradigm of /w/-final verb roots.

In MD, /i/ undergoes lowering when it is not preceded by consonants, and realizes as [e]. When /i/ is in a post-consonantal position, lowering does not occur (/ki/ ‘wood’ is realized as [ki] not *[ke]).

The following SJ-MD correspondences illustrate lowering. The data indicates that lowering occurs not only in word-initial vowels but also in [...Vi...] sequences.

(3)		SJ	MD	
	‘breath’	iki	egi	Native
	‘dog’	inu	enu	
	‘now’	ima	ema	
	‘cold.PRES’	samui	samue	
	‘retreat’	taikjaku	taekjagu	Sino-Japanese
	‘water bottle’	suito:	suetto:	
	‘hiking’	haikinngu	haekinngu	Foreign

¹ The consonant inventory of MD is the same as that of SJ, but the phonotactics differs. The most outstanding features are the banning of /t/ and /k/ in intervocalic position (*kagado* ‘heel’, cf. *kakato* in SJ) and the distribution of [p] in non-geminate environment (*kapto* ‘helmet’, cf. *kabuto* in SJ), etc. MD is also characterized by its rich case-marking system, with three adnominal cases, animacy sensitive case distinction in accusative and dative, and a special case particle for oblique experiencer. Concerning more details, see Miyajima (1961) and Sasaki (1997; 2001).

² Miyajima (1961) described the phonetic realization of these vowel phonemes as follows. The vowels /a, o/ have the same phonetic quality as in SJ, i.e., /a/ is a low vowel and /o/ is a mid back rounded vowel. /u/ is a high back unrounded vowel, slightly front in comparison to SJ /u/. /i/ is realized more central and lower than SJ [i] but distinguished from /u/ with respect to backness, relatively front in comparison to the high central vowel in Tohoku dialects. The MD /e/ is pronounced higher than SJ [e].

³ The palatal glide /j/ is distinguished from SJ [j] by its lower tongue height (determined by the accompanying /i/). /w/ is a velar glide.

Lowering can also be observed in verb stem formation where it appears as the i-e alternation of the thematic vowel. The thematic vowel for the adverbial form varies between [i] and [e]. It appears as [i] in post-consonantal position as shown in the example [tor-i] ‘take-ADV’, and as [e] in positions where it is not preceded by a vowel, as shown in the example [su-e] ‘suck-ADV’. The thematic vowel for the conditional form is realized as [e] not only in post-consonantal position but also in the non-post-consonantal position (see [tor-e] and [su-e]). This suggests that the i-e alternation in the verbal paradigm is better analyzed as a case of lowering (i→e) rather than raising (e→i).

(4) Verb stem formation (partial)

/tor-/ ‘take’	/suw-/ ‘suck, smoke’		
tor-u	su-u	present	
tot-ta	sut-ta	past	
tor-a ne	suw-a ne	negation	
tor-i naŋara	su-e naŋara	adverbial	(i-e alternation)
tor-e ba	su-e ba	conditional	

This dialect has two processes yielding [i] in contexts with no preceding consonants. The relevant processes are /w/-deletion in the verbal paradigm and coalescence applied to /ju/ sequences. The paradigm above indicates that /w/ in root final position drops when it was followed by the vowels other than /a/.

The coalescence is observed in the SJ-MD correspondences below. There are no phenomena like the alternations in verb formation observed in the case of lowering to support the existence of the process. Coalescence occurs irrespective of the presence of preceding consonants, unlike lowering.⁴

(5)

	SJ	MD	
‘mutual aid’	jui	i (iʃiŋodo)	Native
‘citron’	juzu	izu	
‘hot water’	ju	i	
‘operation’	ʃuʒutsu	ʃiʒizu	Sino-Japanese
‘milk’	gju:nju:	gi:ni:	
‘high school’	tʃu:ŋaku	tʃi:ŋagu	
‘post office’	ju:biŋkjoku	i:biŋkjogu	
‘fuse’	hju:zu	hi:zu	Foreign

Both processes can be observed in words such as *eʃʃi:kan* ‘one week’ (*iʃʃu:kan* in SJ) and *ʃikudae* ‘homework’ (*ʃukudai* in SJ).

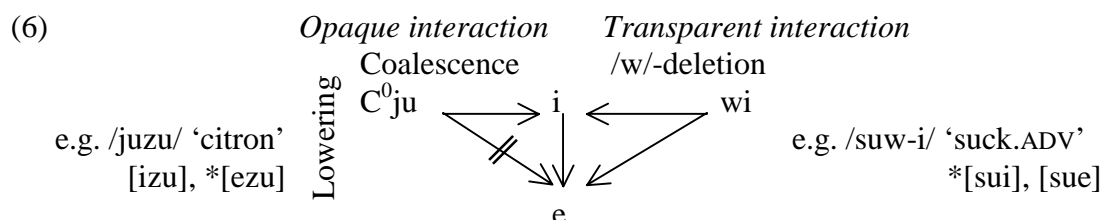
The three phonological processes differ with respect to the context in which they occur. Coalescence is found in SJ-MD correspondences, while /w/-deletion appears in the verb conjugation. Lowering is found in both. The synchronically active status of lowering and /w/-deletion is made clear by the alternations in the verbal paradigm, but there are no morphophonological alternations supporting the synchronic status of coalescence. The situation in the first half of 20th century is important for considering the status of coalescence and the way in which the three processes interact.

The earliest description of the dialect in this area is found in *On’inchosahyo*, published in 1905, based on research conducted in the 1900’s. Lowering and coalescence were already described as characteristics of the dialect therein. This point is maintained also in Kindaichi (1933) and Miyajima (1961). As shown above, coalescence and

⁴ [ʃ] is analyzed as /sj/ phonologically in MD as well as SJ.

lowering are not found only in native and Sino-Japanese vocabulary but also in the foreign vocabulary. Almost all the foreign vocabulary in this dialect is adopted through SJ. Loanwords such as *hju:zu* ‘fuse’ and *haikiŋgu* ‘hiking’ in SJ were introduced after the first research (*hju:zu* 1925, *haikiŋgu* 1930, according to *Shogakkan Nihon Kokugo Daijiten*). The adoption of these loanwords into MD vocabulary must have taken place later. Loanwords undergo modification of their phonetic shapes in accordance with the phonotactic constraints of the target language. The fact that relatively recent loanwords undergo coalescence (*hju:zu* → *hi:zu*) and lowering (*haikiŋgu* → *haekiŋgu*) indicates the active status of these processes at the time of adoption. Thus, the three processes, namely coalescence, lowering, and /w/-deletion, appear to have been active at least until the first half of the 20th century.⁵ The data from my consultants, who were born and grew in the period between 1920’s and 1930’s, reflect this state, where [i] derived from /w/-deletion undergoes lowering while [i] derived from coalescence does not.

The situation described above suggests that the interactions among the three processes are of two types. The interaction between /w/-deletion and lowering is transparent in that the former feeds the later. On the other hand, lowering and coalescence interact opaquely, i.e., the former counterfeeds the later. The situation is illustrated in (6).⁶



The task for us is to provide an explanation for the situation where counterfeeding interaction and feeding interaction co-exist. In the following sections, we will examine what type of theoretical extension is appropriate for the phonological interactions related to lowering in MD.

3. The problem

This section introduces OT-based formulations for each phonological process and clarifies the problem of their interactions.

3.1. Constraints for the respective processes

Generalizations concerning morphophonological alternation or realization of allophones were expressed as rules of the form $A \rightarrow B/X_Y$ in pre-OT analyses. On the other hand, even in the pre-OT analyses, modifications of phonetic shape in borrowings from other

⁵ The dialectal pronunciation [i:ɕiro:] for Yuuzi-roo Ishihara ([ju:ɕiro: iɕihara] in SJ), a movie star who acted around 1950’s–1980’s, indicates the active status of coalescence in the post-World War II period. And the fact that [i:] in Yuuzi-roo does not undergo lowering while the underlying /i:/ in ‘good-PRES’ does (pronounced as [e:]) suggests that the counterfeeding relation between coalescence and lowering obtains.

⁶ The underlying status of the /ju/ sequence may be controversial. It guarantees the presence of [i] that does not undergo lowering. As far as the native and Sino-Japanese vocabulary from the old period is concerned, this leads to a circular argument and absolute neutralization even in the automatic form. But the phonological modification in the recent neologisms mentioned above provides an independent argument for the active status of coalescence and the underlying status of /ju/ sequences.

languages tend to be described as a result of constraints rather than rules, e.g. Shibatani (1973) proposed an account with a surface phonetic constraint for vowel epenthesis in loanwords in SJ. In OT, all phonological phenomena are regarded as a consequence of constraint interactions in a certain constraint ranking. This subsection deals with the constraints responsible for each phonological process and their rankings.

The constraints and the ranking responsible for lowering are presented in (7).

(7) *Lowering* ($i \rightarrow e$):

*i: Avoid /i/ without preceding consonant.

Id(Hi): Specification of [+/- high] must be the same between the Input and the Output.

*i >> Id(Hi)

The constraint *i is a phonotactic markedness constraint which accounts for lowering. The ranking in (7) evaluates the candidate undergoing lowering ($inu \rightarrow enu$), which violates Id(Hi) and satisfies *i, as more harmonic than the faithful candidate ($inu \rightarrow inu$). We must add some words concerning the ranking for lowering. Avoidance of the violation of *i does not result in deletion of [i] and insertion of consonants in front of [i]. This suggests the undominated status of Max- μ and Dep. These faithfulness constraints prohibit the deletion of syllabic elements and segmental insertion, respectively. Backing ($i \rightarrow u$) is also an unavailable option. This must be due to the undominated status of Max(-bk). These faithfulness constraints are not relevant for the opaque and transparent interactions of phonological processes, but they are important for restricting the strategy for *ju avoidance.

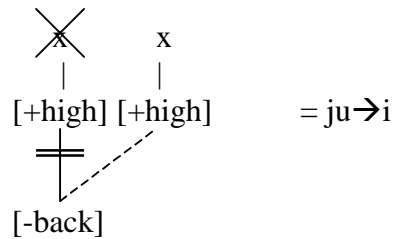
(8) *Coalescence* ($ju \rightarrow i$):

*ju: Avoid the sequence ju.

Id(Bk): Specification of [+/- back] must be the same between the Input and the Output.

Max: Segmental deletion is prohibited.

*ju >> Id(Bk), Max



The coalescence $ju \rightarrow i$ can be analyzed as the spreading of [-back] from /j/ to /u/ and the deletion of the skeletal slot associated with /j/. The phenomenon can be regarded as a consequence of the satisfaction of the phonotactic markedness constraint *ju and the violation of Id(Bk) and Max. The undominated status of Max(-bk) prohibits the simple deletion of [j], which bears [-back] specification. Max(-bk) requires the feature [-back] must remain even when the skeletal slot associated with it is deleted. This forces [-back] spreading to the following vowel. The replacement of [j] with other consonants, which incurs violation of both Max and Dep, is ruled out by the undominated status of Dep.

(9) */w/-deletion* ($wV_{[-low]} \rightarrow V_{[-low]}$)

*wV_[-low]: Avoid the sequence /w/ + non-low vowel.

*wV_[-low] >> Max

In this dialect, as well as in SJ, the only vowel which can follow the glide /w/ is /a/ as mentioned above. The constraint responsible for this distributional restriction is *wV_[-low]. For the input /...wi.../, the candidate [...i...], which undergoes /w/-deletion, is evaluated as more harmonic than the faithful candidate [...wi...] under the ranking in (9). The undominated status of Max- μ bans the candidate undergoing [i] deletion.

3.2. Problem with counterfeeding opacity

The transparent interaction between lowering and /w/-deletion is expected under the ranking in (10), which is a combination of the partial rankings in (7) and (9). This is

illustrated in the Tableau 1. The candidate (c), which satisfies every markedness constraint, is regarded as the most harmonic. This evaluation is compatible with the actual data.

(10) $*wV_{[-low]} \gg *i \gg Id(Hi)$, Max

Tableau 1: *Transparent interaction between lowering and /w/-deletion*

	/suw-i/	$*wV_{[-low]}$	$*i$	Id(Hi)	Max
Faithful	a. suwi	*!			
	b. sui		*!		*
Transparent	c. sue			*	*

The problem arises when we consider the interaction between lowering and coalescence. The evaluation under the ranking (11), a combination of the partial rankings in (7) and (8), is illustrated in Tableau 2, where the candidate (c), which satisfies both $*i$ and $*ju$, is wrongly evaluated as the most harmonic whereas the actual form (b) is treated as sub-optimal.

(11) $*ju \gg *i \gg Id(Hi)$, Id(Bk), Max

Tableau 2: *Failed evaluation of the interaction of lowering and coalescence*

	/juzu/	$*ju$	$*i$	Id(Hi)	Id(Bk)	Max
Faithful	a. juzu	*!				
Opaque (actual)	b. izu		*!		*	*
Transparent	c. ezu			*	*	*

The counterfeeding interaction in MD poses a problem for the optimality theoretic analysis. Two main theoretical extensions have been proposed in the OT literature in order to cope with the problem of counterfeeding opacity within the parallelist approach. One is Local Conjunction (Smolensky 1995) and the other is Comparative Markedness (McCarthy 2002). The rest of this paper will consider the relative merits of the two approaches with respect to the MD data. The best approach will have to account not only for the opaque interaction but also for the transparent interaction. I will examine a Local Conjunction-based account first and an account with Comparative Markedness next.

4. Solution with Local Conjunction

Local Conjunction is a mechanism deriving an undominated constraint on the basis of two lower-ranked constraints. The ranking in (12) is a general schema for counterfeeding opacity proposed by Moreton & Smolensky (2002), where $*x$ and $*y$ stand for the markedness constraints involved and F_1 and F_2 are the faithfulness constraints relevant for the processes.

(12) $*x, F_1 \& F_2 \gg *y \gg F_1, F_2$

In order to account for the MD data, I propose the constraint ranking in (13).

(13) $[Id(Hi) \& Id(Bk)]_{seg}, *ju, *wV \gg *i \gg Id(Hi), Id(Bk), Max$

The undominated status of the locally conjoined constraint $[Id(Hi) \& Id(Bk)]_{seg}$, which has the segment for its domain, prohibits a segment from undergoing a change in the specification for both [high] and [back], although it permits changing either the [high] specification (lowering, $i \rightarrow e$) or the [back] specification (coalescence, $ju \rightarrow i$). This locally conjoined faithfulness constraint can be regarded as the source of counterfeeding opacity between lowering and coalescence. Under the assumption that the relevant faithfulness constraints are Id(Bk) and Id(Hi), the violation of Max in *su-i* (/suw-i/) ‘suck.ADV’ is not expected to be a factor prohibiting further modification, i.e., lowering.

Lowering results in the form [su-e]. Lowering of the first vowel of [izu] (/juzu/) ‘citron’ is banned under the same set of assumptions because it incurs multiple violations of Id(Bk) and Id(Hi). Thus, the proposed constraint ranking handles both the opaque cases and the transparent cases without resorting to serialism. The relevant evaluations are illustrated in Tableaux 3-4.

Tableau 3. *Opaque interaction evaluated with the locally conjoined constraint*

	/juzu/	[Id(Hi)&Id(Bk)] _{seg}	*ju	*wV	*i	Id(Hi)	Id(Bk)	Max
Faithful	a. juzu		*!					
Opaque	☞ b. izu				*		*	*
Transparent	c. ezu	*!				*	*	*

Tableau 4. *Transparent interaction evaluated with the locally conjoined constraint*

	/suw-i/	[Id(Hi)&Id(Bk)] _{seg}	*ju	*wV	*i	Id(Hi)	Id(Bk)	Max
Faithful	a. suwi			*!				
Opaque	b. sui				*!			*
Transparent	☞ c. sue					*		*

5. Failure with Comparative Markedness

Comparative Markedness, advocated by McCarthy (2002), is an extension of standard OT, which divides markedness constraints into two classes: ‘old’ markedness (_OM) and ‘new’ markedness constraints (_NM). _OM constraints are relevant only for candidates that include structure shared with the Fully Faithful Candidate (FFC). FFC is the candidate that does not include any faithfulness constraint violations. On the other hand, _NM constraints are constraints prohibiting certain marked structures not included in the FFC. Counterfeeding opacity is argued to involve a constraint ranking of the form [_OM >> Faith >> _NM].

Assuming that there are two types of *i, _O*i and _N*i, and positing the ranking in (14), counterfeeding cases are analyzed correctly. The evaluation for the interaction of lowering and coalescence is illustrated in Tableau 5.

(14) *ju, *wV >> _O*i >> Id(Hi), Id(Bk), Max >> _N*i

Tableau 5. *Opaque interaction evaluated with Comparative Markedness*

	/juzu/	*ju	*wV	_O *i	Id(Hi)	Id(Bk)	Max	_N *i
Faithful	a. juzu	*!						
Opaque	☞ b. izu					*	*	*
Transparent	c. ezu				*	*	*!	

The ‘new’ onsetless *i* derived through coalescence does not incur a violation of the higher-ranked _O*i although it violates the lower-ranked _N*i. The opaque candidate (b) is evaluated as more harmonic than the transparent (but less faithful) candidate (c).

The problem arises on the interaction between lowering and /w/-deletion. Under the constraint ranking in (14), the opaque candidate is evaluated as the most harmonic, as illustrated in Tableau 6. The actual form expected through the transparent interaction is regarded as less harmonic than the opaque one.

Tableau 6. *Failed evaluation with Comparative Markedness*

	/suw-i/	*ju	*wV	_O *i	Id(Hi)	Id(Bk)	Max	_N *i
Faithful	a. suwi		*!					
Opaque	☞ b. sui						*	*
Transparent (actual)	c. sue				*		*!	

Thus, Comparative Markedness cannot offer a solution for dealing with both opaque and transparent interaction. On this approach, the effort of resolution for opacity results in the failure of accounting for the transparent interaction in the respective cases.

6. The source of inadequacy of Comparative Markedness for MD interactions

I have shown that Local Conjunction makes the correct predictions for the counterfeeding opacity and the transparency data in the MD, whereas Comparative Markedness fails to do so. In what follows I will present some remarks on the source of this difference between the two mechanisms.

The key difference between the two proposals lies in the distinction they make among ‘new’ structures that have different sources. Consider the situation where there are two processes (P_1, P_2) which yield a structure (XAY) which meets the conditions for another process ($P_3: A \rightarrow B/X_Y$). It is possible to assume four relationships among P_3 and the other processes. The table in (15) illustrates these potential relationships. Phonological interactions related to lowering in MD fall in the case of (15b) or (15c).

(15)

	P_1	P_2	
a. P_3	transparent	transparent	Totally transparent
b. P_3	transparent	opaque	Partially opaque
c. P_3	opaque	transparent	Partially opaque
d. P_3	opaque	opaque	Totally opaque

In the situations (15b) and (15c), the XAY derived from one process undergoes P_3 while the XAY derived from the other process does not. This distinction between XAY from P_1 and P_2 can be captured through the distinction between faithfulness constraints because the ‘new’ XAYs from the distinct processes are not different except for the violations of faithfulness constraints incurred by each process.

With Local Conjunction, it is possible to distinguish the relevant and irrelevant faithfulness constraints and to put the conjoined constraints consisting of the relevant faithfulness constraints into the undominated position. This covers the partially opaque situation where the locally conjoined faithfulness constraint blocks the further modification to the ‘new’ XAY from one process but permits the further modification to the ‘new’ XAY from the other process.

On the other hand, Comparative Markedness cannot make such distinctions among the ‘new’ XAYs. The XAY derived from P_1 and the XAY from P_2 are equally ‘new’ because XAYs derived through violations of *any* faithfulness constraints incur the ‘new’ markedness constraint, namely $_N^*XAY$ by definition. The ranking where the ‘new’ markedness constraint is dominated by the faithfulness constraints creates a totally opaque situation. Under this ranking XAYs from P_1 and P_2 are equally free from the application of P_3 and P_3 applies only to the underived XAYs. The ranking where the ‘new’ markedness constraint is undominated ensures that the derived XAYs from both P_1 and P_2 undergo P_3 while underived XAY does not undergo P_3 . The ranking where the ‘new’ and ‘old’ markedness constraints are undominated makes the underived XAY and the XAYs from P_1 and P_2 all undergo P_3 . In spite of the difference concerning the underived XAY, Comparative Markedness can describe only two of the four situations of the interactions among P_1, P_2 , and P_3 , namely total opacity and total transparency. It cannot describe the partially opaque situations by itself. Thus, Comparative Markedness makes the wrong prediction when opacity and transparency related to the same process co-exists. What is most important in the case of partial opacity is that the faithfulness constraints violated through P_1 and P_2 are different. This point can be captured through the extension of faithfulness constraints (as illustrated throughout the paper with locally conjoined faithfulness constraints) but not through the extension of markedness (Comparative Markedness is an instance).

7. Concluding remarks

Comparative Markedness can account for counterfeeding opacity itself, but it cannot deal with cases where opacity and transparency co-exist. Local Conjunction can accommodate this type of situation, at least as far as the interaction observed in the MD is concerned. This does not mean, however, that Local Conjunction is omnipotent in all the cases where transparency co-exists with opacity. According to McCarthy (1999), counterfeeding opacity can be classified into two types: counterfeeding opacity on focus and counterfeeding opacity on environment. The MD data is an example of the former type. It seems that Local Conjunction cannot cope with the second type of counterfeeding opacity because the violations of the faithfulness constraints are not within a single domain.

A known case where a counterfeeding interaction on environment co-exists with a transparent interaction comes from Yokuts, where lowering interacts with rounding harmony opaquely while epenthesis and rounding harmony interact transparently. McCarthy (*ibid.*) has proposed an account for the Yokuts data using Sympathy. Sympathy is another extension of the faithfulness constraints. The data cannot be accounted for with Comparative Markedness. The attempt to capture the opaque interaction between lowering and rounding harmony with an ‘old’ markedness constraint for rounding harmony in higher position in the ranking ends in the wrong prediction that the transparent interaction between epenthesis and rounding harmony will be ruled out. Thus, Comparative markedness fails to account for the situation where opacity and transparency co-exist in the case of counterfeeding on environment, as well as on focus.

It seems that the problem of the situation where transparency and opacity co-exist involves more than a comparison between Comparative Markedness and Local Conjunction, and might require a reconsideration of the relation between the markedness extensions and the faithfulness extensions as a whole. In order to find a general answer to this problem other markedness extension proposals, such as Targeted Constraint, need to be included in the investigation.

Appendix

In this paper, we assumed that the coalescence and lowering occur throughout the vocabulary even in the non-derived environments. Under this assumption, coalescence is a kind of automatic absolute neutralization. This type of absolute neutralization is not excluded in the literature because of its usefulness in accounting for some linguistic phenomena (see Kiparsky 1973:67). But, under the strong version of the Alternation Condition (Kiparsky *ibid.*), which excludes any absolute neutralization, the phonological interactions presented above are analysed as the combination of the Derived Environment Effect and counterfeeding opacity. This move of assumption does not touch the conclusion concerning the inapplicability of Comparative Markedness to the problem. This appendix presents some remarks concerning this point.

The strong version of the Alternation Condition requires that every surface [i] and [e] correspond to the underlying /i/ and /e/, respectively, unless they alternate with other segments. This leads to the inactiveness of coalescence and lowering at least in native and Sino-Japanese vocabulary. Under this assumption, lowering is active only in the verbal paradigm, where /w/-deletion provides the environment for lowering. This situation is captured through the constraint ranking with Comparative Markedness in (A1).

(A1) $N^*i, *wV \gg \text{Id-IO}(\text{Hi}), \text{Id-IO}(\text{Bk}) \gg o^*i, *ju$

The partial ranking [_N*i >> Id-IO(Hi) >> o*i] reflects the general schema for Derived Environment Effect [_NM >> Faith >> oM], which assures lowering applies only in derived environments. The higher ranking of Id-IO(Bk) versus *ju guarantees the inactiveness of coalescence.

The fact that lowering and coalescence are found in relatively recent neologisms might be explained if these loanwords are assumed not to acquire the underlying form (i.e., Input) status. Let us assume that what associates the correspondence relation to the Output is the phonetic form in SJ and the faithfulness constraints for them are expressed as Id-SJO(feet). The active status of lowering and coalescence in the neologisms can be analysed as a result of the partial constraint ranking [o*i, *ju >> Id-SJO(Hi), Id-SJO(Bk)]. However, the combination of this partial ranking with the ranking (A1) leads us to the wrong prediction.

(A2) _N*i, *wV >> Id-IO(Hi), Id-IO(Bk) >> o*i, *ju >> Id-SJO(Hi), Id-SJO(Bk)

The ranking in (A2) predicts the transparent interaction between coalescence and lowering for phonetic modifications in neologisms and it predicts that the SJ [ju] will be modified as [e]. In order to accommodate the analysis to the fact, we should posit a locally conjoined constraint [Id-SJO(Hi)&Id-SJO(Bk)] with undominated status, which blocks further modification (i.e., lowering) to the [i] derived from coalescence. Thus, even under the strong version of Alternation Condition, Comparative Markedness cannot provide a correct analysis for the MD interaction. The interaction in MD cannot be handled without Local Conjunction.

References

- Kindaichi, Haruhiko. 1933. Kantôheiya-chihô no on'inbunpu. *Hôgen Kenkyû* 8. 1-46.
- Kiparsky, Paul. 1973. Phonological representations. In Osamu Fujimura (ed.), *Three Dimensions of Linguistic Theory*, 1-136. Tokyo: TEC.
- McCarthy, John. 1999. Sympathy and phonological opacity. *Phonology* 16. 331-399.
- McCarthy, John. 2002. *Comparative Markedness*. Ms., University of Massachusetts.
- Miyajima, Tatsuo. 1961. Hôgen no jittai to hyôjungoka no mondaiten 6: Fukushima, Ibaraki, Tochigi. In Misao Tôjô (ed.), *Hôgengaku Kôza 2: Tôbu Hôgen*, 236-263. Tokyo: Tôkyôdô.
- Moreton, Elliott & Paul Smolensky. 2002. *Typological consequences of Local Constraint Conjunction*. Ms., Johns Hopkins University.
- Sasaki, Kan. 1997. Possessive, genitive and adnominal locative in the Mitsukaido dialect. In Tooru Hayasi & Peri Bhaskararao (eds.), *Studies in Possessive Expressions*, 117-41. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.
- Sasaki, Kan. 2001. The grammatical function of oblique elements in the Mitsukaido dialect of Japanese. In Shigeru Sato and Kaoru Horie (eds.), *Cognitive-Functional Linguistics in an East Asian Context*, 175-205. Tokyo: Kurosio Publishers.
- Shibatani, Masayoshi. 1973. The role of Surface Phonetic Constraints in generative phonology. *Language* 49. 87-106.
- Smolensky, Paul. 1995. On the Internal Structure of the Constraint Component *Con* of UG. Ms., John Hopkins University.

Multi-level Evaluation in Optimality Theory: Evidence from Word-formation and Morpheme Identification*

Hideki Zamma

Kobe City University of Foreign Studies

Abstract. This paper proposes that a distinct stage of morphological evaluation is necessary within Optimality Theory. Under a standard model which recognizes only one 'level' of calculation, several problems remain unsolved. The data which support this claim come from word-formation and morpheme identification in English and Japanese.

1. Introduction

Current standard OT assumes that constraints only evaluate output forms, but not inputs, within a single level. This assumption has succeeded to a great extent in eliminating the notion of classical 'derivation,' which has been criticized for its abstractness. Several studies have also shown that it is indeed possible to analyze many phonological phenomena under this assumption.

Under such an assumption, however, several facts of word-formation cannot be accounted for. As the discussion in Section 2 clarifies, morphological requirements which do not allow any base to violate them cannot be properly analyzed. Moreover, given the general assumptions of Richness of the Base and Lexicon Optimization, identity of morphemes is not truly guaranteed, as shown in Section 3.

In order to solve these problems, I will propose that a distinct level of morphology should be established, so that evaluation is carried out at multiple levels. Section 4 shows how easily this assumption can account for the problems at hand. Section 5 concludes the paper.

2. Null-parse vs. Faith violation

2.1. *The facts of English suffixation*

As shown in various studies, affixes put particular restrictions on the bases to which they attach. The restrictions can be categorized into several types, one of which is exemplified by *-ory* and *-ive*. These are required to attach to bases which end with /s/ or /t/ (cf. Zamma (1994a, 2000)).¹ In (1), the base ends with a segment required by the suffixes, and thus simple suffixation takes place.

* I am grateful to Mark Campana for suggesting stylistic improvements, and to Jennifer Spenader for practical help.

¹ This description is simplified in some ways. In fact, *-ory* and *-ive* distinguish single /s, t/ from those in clusters where the preceding segment is homorganic to /s, t/. Thus, *present* takes *-ative* (*presentative*) rather

- (1) a. -ory: dismiss-ory, vomit-ory, excret-ory, deposit-ory, contribut-ory
- b. -ive: reflex-ive, regress-ive, act-ive, effect-ive, prohibit-ive, possess-ive

In (2), on the other hand, the final segment is neither /s/ nor /t/. In these cases, a special suffix *-ate* is introduced between the suffix and the base, whose final segment clearly satisfies the requirement.²

- (2) a. -atory: sign-atory, reform-atory, observ-atory, declar-atory, inflamm-atory
- b. -ative: accus-ative, conserv-ative, provoc-ative, compar-ative, affirm-ative

Sporadically, the final segment of the base is changed so that the requirement of the suffixes is satisfied.

- (3) a. /d/ → /s/: expansive (< expand), decisive (< decide), abrasive (< abrade)
- b. /z/ → /s/: abusive (< abuse), effusive (< effuse)
- c. /r/ → /s/: cohesive (< cohere), adhesive (< adhere)
- d. /ʃ/ → /t/: admonitory (< admonish), punitory (< punish)

These facts suggest that the suffixal requirement can be satisfied by modifying the original input in some ways.

A similar phenomenon is observed with *-al*, which imposes the opposite requirement to *-ory/-ive*: the base-final segment should not be /s/ or /t/. If it is /s/ and /t/, /i/ and /u/ are inserted before the suffix respectively. Compare (4a) with (4b, c).

- (4) a. verb-al, physic-al, economic-al, prim-al, origin-al, person-al, adjectiv-al
- b. fac-ial, rac-ial, offic-ial, sacrific-ial
- c. act-ual, intellect-ual, habit-ual, spirit-ual

The suffix *-en* has a completely different type of requirement, whereby the suffix is required to attach to monosyllabic bases which end with an obstruent (cf. Halle (1973)).³ In this case, an output form is never produced when the requirement is violated (5b).

- (5) a. tight-en, loos-en, stiff-en, weak-en, wid-en, deep-en, length-en
- b. *green-en, *blue-en, *tall-en, *clear-en, *narrow-en, *complex-en

Similarly, *-ize* does not allow any output which violates its requirement, whereby the base must not have final stress (cf. Raffelsiefen (1996)).

- (6) a. rándom/rándamize, sálmon/sálmonize, fóreign/fóreinize, síster/sísterize
- b. ápt/*aptize, fírm/*fírmize, corrúpt/*corruptize, obscéne/*obscenize

The bases in (6b), which have their primary stress on the final syllable, do not have forms with *-ize*, contrary to (6a).

These facts are in clear contrast to the cases of *-ory*, *-ive* and *-al*. Recall that words with these suffixes modify the base so that an output will satisfy the suffixal requirement. *-en* and *-ize* never allow any output when the input string will violate the requirement. Keeping this contrast in mind, let us consider how these facts can be analyzed within Optimality Theory

2.2. An OT analysis and its problem

The first type of requirement, which is represented here by *-ory*, can be easily accounted for. Let us assume the constraints in (7) and their ranking in (8):

- (7) a. **Align(-ory/-ive, L, /s, t/):** -ory and -ive must attach to bases which end with /s/ or /t/.

than *-ive*. Moreover, the /s/ in the cluster /ns/ behaves in the same way as a single /s/ (e.g. *offensive* < *offence*). See Zamma (2000) for details.

² As discussed in Zamma (1994b), the *-at-* in *-atory/-ative* should be regarded as the suffix *-ate* in order to account for stress behavior and vowel length in the words containing them.

³ This description is also simplified: *-en* cannot be attached to an obstruent when it is in certain clusters. See Halle (1973) for details.

- b. **Faith(Base)**: The base should not be modified.
 c. **Faith(-ate)**: *-ate* should not be inserted or deleted.

(8) **Align(-ory), Faith(Base) » Faith(-ate)**

The constraint (7a) is a requirement of the suffix. The Faithfulness constraints (7b-c) militate against modification of the relevant morpheme. Of these, (7c) is ranked lowest.

Next, observe how these constraints and their ranking can produce the correct output. The tableau (9a) is for bases which satisfy the suffixal requirements, and the tableau (9b) for bases which do not.

(9) a.

dismiss + -ory	Align(-ory)	Faith(B)	Faith(-ate)
☞ dismiss-ory			
dismitt-ory		*!	
dismiss-atory			*!
Ø		*!*****	

b.

sign + -ory	Align(-ory)	Faith(B)	Faith(-ate)
sign-ory	*!		
sigt-ory		*!	
☞ sign-atory			*
Ø		*!****	

In (9a), the form in which the suffix is simply attached violates none of the constraints and is thus selected as optimal. In (9b), on the other hand, such simple suffixation violates the top-ranked constraint of the suffixal requirements. Neither modification of the base-final segment nor null-parsing is a good solution compared to insertion of a special suffix *-ate* -- hence the third candidate wins.⁴ As to the words in (3), the ranking between **Faith(B)** and **Faith(-ate)** is reversed: thus a part of the base is modified instead of introducing *-ate*. The case of *-ive* can be similarly accounted for. In sum, resolution of the first type of violation is determined via the relative ranking of Faithfulness constraints.

The analysis of the other type, however, raises a crucial problem for the current architecture of Optimality Theory. Recall that this type of requirement produces no output when the input sequence violates it. Let us review here how these phenomena are treated in the literature. Prince and Smolensky (1993:49) and McCarthy and Prince (1993:112) both analyze these with the following schema of constraint ranking:

(10) **Markedness » MParse**

MParse is a constraint which requires an input to have an output, whose definition is given below:

(11) **MParse**: Morphemes are parsed into morphological constituents.

Markedness, on the other hand, is a general term for any kind of constraint which forces phonological change on the input. Given the ranking in (10), they claim that null-parsing is selected as the optimal candidate.

(12)

A + x	Markedness	MParse
A - x	*!	
☞ Ø		*

As shown in (12), the null-parse candidate is more optimal than the simply-affixed one, which violates **Markedness**.

⁴ A violation of **Faith(Base)** is calculated here by the number of segments which are not faithful to the input. Other calculations bear similar results, aside from the number of violation marks.

At a first glance, this seems to work. If one considers the analysis more deeply, however, it turns out to have a fatal problem. Let us discuss the case of *-en*-suffixation. First, we will take a case in which the input has an actual form. By observing the distribution of *-en*, it is appropriate to assume the following **MorphReq** constraints for this suffix:

(13) a. **Base** ≤ **σ**: The base should be monosyllabic.

b. **Align(-en, L, [-son])**: *-en* must attach to bases which end with an obstruent.

The actual outputs can be predicted in the schema given above, where **Markedness** -- in this case **MorphReq** -- is ranked above **MParse**.

(14)

tight + -en	MorphReq	MParse	Faith(B)
☞ tight-en			
tighd-en			*!
∅		*!	***

The null-parse candidate is never produced in this case. It is important to note here that **Faith(Base)** is necessary somewhere in the hierarchy so that modified outputs can be eliminated (as exemplified by the second candidate). Although we temporarily put it lowermost in (14), the same result is produced when it is placed higher.

In cases where the null-parse candidate **MUST** be selected, however, the schema does not work successfully. Consider the unattested form **greenen*.

(15) a.

green + -en	MorphReq	MParse	Faith(B)
green-en	*!		
unwanted output ☞ greet-en			*
∅		*!	****

b.

green + -en	MorphReq	Faith(B)	MParse
green-en	*!		
unwanted output ☞ greet-en		*	
∅		***!	*

Regardless of whether **Faith(Base)** is ranked lowest (15a) or between **MorphReq** and **MParse** (15b), the schema predicts an unwanted output in the middle. Recall that **Faith(Base)** must be present somewhere in order to eliminate the unwanted candidate of the actual form. The faithfulness constraint also eliminates the null-parse candidate in favor of one which minimally modifies the base, because null-parsing incurs a violation for each of the segments contained in the input.

The highest ranking of **Faith(Base)** does not predict the correct form either. Such a ranking predicts that the input will be parsed without any modification.

In sum, the morphological requirements of suffixes show differences in strength -- one satisfied by modification of the base, and the other not producing an output at all. This contrast is problematic for the current OT, because it predicts only the former case. In the current OT where only outputs are evaluated, an output can satisfy a Markedness constraint by modifying the input, as in the cases of *-ory* and *-ive*, as long as it is ranked higher in the hierarchy than Faithfulness.

3. Morpheme identification

3.1. Japanese palatalization and a problem in OT

Another crucial problem arises under the current architecture of OT: the identification of morphemes. In this section, we will see examples from Japanese and English.

It is well known that actual words in Japanese must in principle end with a vowel -- except for the moraic nasal (cf. Itô (1986) among others). Japanese verb stems, however, can end either with a vowel (16a) or a consonant (16b).

- (16)
- | | | |
|----------------|-------------------|---------------|
| | <i>indicative</i> | <i>polite</i> |
| a. mi- 'look' | mi-ru | mi-masu |
| b. yom- 'read' | yom-u | yom-i-masu |

Because inflectional suffixes always follow stems, consonant-final stems can appear to satisfy the restriction of open-syllabicity on the surface, but there is no case in which the stem appears as is, i.e. without any suffix (e.g. **yom*). In other words, consonant-final stems are bound.

On the other hand, there is a palatalization rule in Japanese, which turns non-labial consonants into palatals before the vowel /i/. Thus, some forms of consonant-final verb stems appear with the stem-final consonant palatalized, as shown below:

- (17)
- | | | |
|-----------------|---------------------|---------------------------|
| a. kak- 'write' | kak-u | kak ^J -i-masu |
| b. kag- 'sniff' | kag-u | kag ^J -i-masu |
| c. kas- 'lend' | kas-u | ka ^{sà} -i-masu |
| d. kat- 'win' | kats-u ⁵ | ka ^{cà} -i-masu |
| e. sàin- 'die' | sàin-u | sàin [□] -i-masu |

This phenomenon can be captured in current OT. Let us tentatively assume that the following constraints are ranked as in (19).

- (18) a. **Palatalization**: A non-labial consonants palatalize in front of /i/.
 b. ***CJ**: Palatalized consonants are not allowed.

- (19) **Palatalization** » ***CJ** » **Faith**

The constraint in (18b) is necessary because palatalized consonants are not allowed except in front of /i/. With the ranking (19), the (non-)palatalization of /k/ in indicative and polite forms is correctly predicted.

- (20) a.

kak + -u	Pal	*CJ	Faith
☞ kak-u			
kak ^J -u		*!	*

- b.

kak + -i + -masu	Pal	*CJ	Faith
kak-i-masu	*!		
☞ kak ^J -i-masu		*	*

When a suffix beginning with a vowel other than /i/ follows the base, no change occurs as in (20a). Only in cases where an /i/-initial suffix follows does the stem-final consonant palatalizes.

A crucial problem in the identification of the verb stem arises here, given the general assumptions of Richness of the Base and Lexicon Optimization. Under the former, the input form of the base can be either /kak-/ or /kak^J-/ both for [kak-u] and [kak^J-i-masu], because Gen can produce the actual forms in either case. The tableaux below show that inputs with stem-final /k^J/ can successfully produce the correct outputs.

⁵ /t/ undergoes spirantization in front of /u/. The underlying /t/ can be identified in other forms such as the imperative [kat-e].

(21) a.

kak J + -u	Pal	*C J	Faith
☞ kak-u			*
kak J -u		*!	

b.

kak J + -i + - masu	Pal	*C J	Faith
kak-i-masu	*!		*
☞ kak J -i-masu		*	

In order to determine a unique input, a device called Lexicon Optimization is postulated in Optimality Theory, through which an input whose optimal output incurs the least number of violations is selected as optimal. Let us apply it to the case at hand.

(22) a.

input	output	Pal	*C J	Faith
☞ kak-u	kak-u			
kak J -u	kak-u			*!

b.

input	output	Pal	*C J	Faith
kak-i-masu	kak J -i-masu		*	*!
☞ kak J -i-masu	kak J -i-masu		*	

The optimal outputs in (20) and (21) are compared with respect to their constraint violations. As shown in (22), the optimal inputs selected for [kak-u] and [kak**J**-i-masu] are both faithful -- /kak-u/ and /kak**J**-i-masu/ -- since they incur fewer violations accordingly.

Now, the problem arises: different inputs are selected for each form, i.e. /kak-/ for indicative and /kak**J**-/ for polite. In such a case, how can /kak-/ and /kak**J**-/ be identified as the same morpheme? It is impossible to maintain the identity of the stem in a verbal conjugation: indicative and polite forms will be regarded as different words, having distinct forms as their input.⁶

3.2. English Cluster Simplification

A similar problem arises in English Cluster Simplification. English has a famous rule which deletes an unsyllabifiable segment, as in (23).

- (23) a. sign [sain] signature [sign□tə□r]
 b. hymn [him] hymnal [himn□l]
 c. bomb [b**A**m] bombard [b**A**mb**A**□rd]

⁶ Actually, Prince and Smolensky (1993) suggest a solution to this kind of problem by introducing the following constraint into the grammar (p.196):

- (i) ***Spec**: Underlying materials must be absent.

When this constraint is ranked higher than **Faith**, an analysis containing fewer morphemes is more optimal than, and thus wins over, one which has several allomorphs per morpheme. In the relevant case, having two allomorphs *kak-* (for indicative) and *kak**J**-* (for polite) is worse than having just *kak-* for both forms. The input of the stem is thus uniquely determined by this ranking. This analysis, however, is improper particularly in that ***Spec** evaluates input forms directly, even though current OT assumes that constraints evaluate only output forms.

These facts can be analysed tentatively as in (25), making use of the constraint in (24) (N.B.: periods indicate syllable boundaries).⁷

(24) **Sonority Sequencing Principle:**

Syllables must be made according to the Sonority Hierarchy.

(25) a.

sign	SSP	Faith
.sign.	*!	
☞ .sain.		*

b.

sign + -ature	SSP	Faith
☞ .sig.n□.tā□r.		
.sai.n□.tā□r.		*!

When no suffix follows the base, the least sonorant segment (i.e. /g/ in *sign*, /b/ in *bomb*, etc.) is deleted in order to satisfy the **Sonority Sequencing Principle** (25a). When a vowel-initial suffix follows the base, on the other hand, the segment in question can be syllabified and thus deletion does not apply (25b).

Given Richness of the Base, /sain/ is also a possible input in unsuffixed form. Of course, such an input produces the correct output.⁸

(26)

sain	SSP	Faith
.sign.	*!	*
☞ .sain.		

Lexicon Optimization selects /sain/ as the input for [sain] as it incurs the fewest number of violations.

(27)

input	output	SSP	Faith
sign	.sain.		*
☞ sain	.sain.		

On the other hand, the input for [sign□tā□r] would be /sign□tā□r/ (cf. fn.8). Again, Lexicon Optimization selects different forms as the input for each of these words; i.e. /sain/ for [sain] and /sign/ for [sign□tā□r]. How then can we guarantee that these two words are morphologically related, being comprised of the same morpheme?

In sum, this problem arises wherever the base form does not appear as it is; that is, when it is a bound morpheme (as in the case of Japanese inflection), or when it is modified (as in the case of English cluster simplification). When the base appears in the same form as its 'underlying' structure, its identity can be maintained among morphologically-related words via Output-Output Correspondence. This is impossible in the cases we have seen here, because the base form itself, to which the derived words may refer, is modified. As long as the assumptions of Richness of the Base and Lexicon Optimization are maintained, the base and its derived word must be completely distinct -- morphologically unrelated.

⁷ Additional constraints are of course necessary, for example, to guarantee the vowel lengthening in (23a).

⁸ In the case of suffixed forms, it is hard to ensure that a particular segment which is absent in the input is inserted in front of suffixes (e.g. /g/ for /sain/, /b/ for /bAm/, etc.).

4. An alternative analysis

4.1. A proposal

Let us now summarize the problems we face. First, requirements that never produce an actual output form, as observed in the English word-formation facts, cannot be accommodated within the standard framework: there should be an output which satisfies the requirement as long as it is imposed on the output. On the other hand, identification of morphemes is not guaranteed for Japanese verbal inflection or English Cluster Simplification: Lexicon Optimization cannot select a consistent input between the base and the derived words. These two problems both arise from the standard architecture of current Optimality Theory, where there are no restrictions on input forms, and only outputs are evaluated at a single level.

(28) Architecture of current OT:

input → Gen → candidates → H-Eval → output

The problems are easily resolved if we assume a distinct component of morphology, where inputs are identified and combinations of morphemes are evaluated. Legitimate inputs are then sent to the phonological component.

(29) An Alternative Approach (cf. Zamma (1997), etc.):

Input Formation → {Gen, H-Eval} → Phonetics

This is to say that a distinct component of morphology is available before phonology, where morphemes are created and combined. In other words, morphemes and their combination are realized at the morphological component before being evaluated in the phonological component. A similar approach is in fact given in Harmonic Phonology (cf. Goldsmith (1993)).

(30) Morphophonemics → Phonology → Phonetics
Morphology

Note also that similar multi-level approaches have been proposed within Optimality Theory by Booij (1997), Kaun (1998), Kiparsky (2000), etc. Moreover, recent work within Dispersion Theory (cf. Flemming (1995), Padgett (1997), etc.) can be regarded as a computation of the morphological component, as this is a theory that determines the phoneme inventory of a language, that is, the elements with which morphemes are created. These studies suggest that the current approach is reasonable within the framework of Optimality Theory.

Moreover, the model proposed here is not reminiscent of classical derivation, which Optimality Theory dispensed with. Note that it is just a division of grammatical components: morphological, phonological and phonetic. Below we will see how this proposal can properly accommodate the cases seen above.

4.2. Morpheme identification

Let us first consider the problem of morpheme identification. In the proposed model, each morpheme is created in the morphological component by arranging the possible phonemes of the relevant language. At this stage, Japanese /kak-/ and English /sign/ are created. Next in the phonological component, palatalization alters the stem-final segment before the suffix /i/ in Japanese, while Cluster Simplification in English deletes the less-sonorant segment /g/ and lengthens the preceding vowel when no vowel follows, producing [sain].

(31) Morphophonemics	→	Phonology (Gen + Eval)
identification of morphemes		palatalization, cluster simplification
kak-, sign, etc.		(= (20b), (25a), etc.)

In other words, every morpheme is identified uniquely at the morphological component. Consequently, the problem of morpheme identification never arises in this alternative model.

4.3. Null-parsing

Recall that the problem of null-parsing arises from the assumption that constraints only evaluate output forms. As long as there is an input, any output can be produced by Gen, and an output can satisfy a Markedness constraint via minimal modification of the input form at the cost of a **Faith** violation. This is inevitable in the current standard architecture of Optimality Theory.

The fact that a strong requirement never allows such modification suggests that it is a restriction of quite a different nature from the common output constraints. It must be the input that is restricted by this requirement, and thus some possible combinations of morphemes are completely disallowed from having an actual output form. In other words, an unattested form is absent as an input to phonology from the start: a form which is absent from the input will not naturally appear in the output.

As we argued in Section 2.2, it is possible to satisfy a constraint *phonologically* by modifying the input sequence of sounds. If a sequence of morphemes is totally absent, it is considered that the combination of the morphemes is *morphologically* disallowed. Thus, in the model (29), an unattested form is ruled out at the morphological component before it is sent to the phonological component.

Let us see how the *-en* case is resolved. First in the morphological component, it is evaluated as to whether a given combination of morphemes is legitimate or not. Assuming that OT architecture is also present in the morphology, the evaluation goes as follows:

(32) Morphology

green + -en	MorphReq	MParse
green-en	*!	
☞ Ø		*

MorphReq, which as its name suggests naturally applies at the morphological level, eliminates the candidate **green-en*. Consequently, the null-parse candidate wins out. Note that, unlike the phonological component, the morphological component does not seem to have Faithfulness constraints. This is natural when we note the fact that the meaning of a word is never modified in order to satisfy a semantic constraint on word-formation. Morphemes are either present as they are, or else, absent: thus **MParse**.

Actual forms, on the other hand, are produced in the following way:

(33) a. Morphology

tight + -en	MorphReq	MParse
☞ tight-en		
Ø		*!

b. Phonology

tight-en	Markedness	Faith(B)
☞ tight-en		
tighd-en		*!
Ø		*!***

First in the morphological component, the combination of *tight* plus *-en* is evaluated and the combined form is selected over null-parsing. Next, the sequence is evaluated in the phonological component: in this case the sequence of phonemes /taɪt□n/ does not violate

any of the **Markedness** constraints, and thus is selected as optimal without any alternation.

Moreover, the case of *-ory* can be analyzed properly. Since the restriction applies at the phonological component, any combination of the suffix with the base passes through the morphological component.⁹ At the phonological component, where the restriction in (7a) applies, those combinations which violate the restriction modify the base in order to satisfy it, as shown in the evaluation in (9).

5. Conclusion

Under the current standard model of OT, null-parsing cannot be properly accounted for, although some have claimed that it can. Such researchers say that **M**Parse is a constraint that guarantees null-parsing, yet actually it has no role in selecting null-parse candidates as optimal, because such candidate always loses out to those which incur minimal violations of **Faith**. Rather, **M**Parse should apply at a distinct level of morphology, where certain combinations of morphemes are ruled out (that is, null-parsed).

Moreover, Richness of the Base together with Lexicon Optimization cannot guarantee the identity of those morphemes which do not appear independently on the surface. Lexicon Optimization selects different forms for such stems from their derivatives. This problem also arises under the current architecture with single-level evaluation, where there is no stage at which a morpheme is identified.

Taken together, these facts strongly support the claim of multi-level evaluation within Optimality Theory. As discussed in this paper, postulation of a distinct morphological component easily resolves the problems at hand. Although the model proposed in this paper needs further refinement, it is likely that this approach will produce worthwhile results.

References

- Booij, G. 1997. Non-derivational Phonology Meets Lexical Phonology. In I. Roca (ed.), *Derivation and Constraints in Phonology* (New York: Oxford Univ. Pr.), 261-288.
- Flemming, E. 1995. *Auditory Representations in Phonology*. Doctoral dissertation, University of California, Los Angeles.
- Goldsmith, J. 1993. Harmonic Phonology. In J. Goldsmith (ed.), *The Last Phonological Rule* (Chicago: the Univ. of Chicago Pr.), 21-60.
- Halle, M. 1973. Prolegomena to a Theory of Word-formation. *Linguistic Inquiry* 4, 3-16.
- Itô, J. 1986. *Syllable Theory in Prosodic Phonology*. Doctoral dissertation, University of Massachusetts, Amherst.
- Kaun, A. 1998. Input Constraints in Tamil. *CLS* 34:2, 79-93.
- Kiparsky, P. 2000. Opacity and Cyclicity. *The Linguistic Review* 17, 351-365.
- McCarthy, J. and A. Prince. 1993. Prosodic Morphology I: Constraint Interaction and Satisfaction. Ms., University of Massachusetts, Amherst and Rutgers University
- Padgett, J. 1997. Perceptual Distance of Contrast: Vowel Height and Nasality. *Phonology at Santa Cruz* 5, 63-78.
- Prince, A. and P. Smolensky. 1993. Optimality Theory: Constraint Interaction in Generative Grammar. Ms., Rutgers University and University of Colorado.
- Raffelsiefen, R. 1996. Gaps in Word Formation. In U. Kleinhenz (ed.), *Interfaces in Phonology* (Berlin: Akademie Verlag), 194-209.
- Zamma, H. 1994a. Phonological Requirements on Suffixation. *Tsukuba English Studies* 13, 21-41, University of Tsukuba.

⁹ Some combinations are also ruled out by semantic restrictions.

- Zamma, H. 1994b. Accentuation of *-ory*, *-ive*, and *-ion*. *English Linguistics* 12, 248-271.
- Zamma, H. 1997. How Are Inputs Generated in Optimality Theory? *Tsukuba English Studies* 16, 1-20, University of Tsukuba.
- Zamma, H. 2000. Stricture and Word Formation in English (Review Article: *Stricture in Feature Geometry*, by J. Padgett, CSLI Publications, Stanford, CA, 1995). *English Linguistics* 17:2, 573-590.

OT variations in syntax

Local vs. global optimization in syntax: a case study

Gereon Müller

IDS Mannheim

Abstract. The main goal of this paper is to argue for an approach to optimization in syntax that is not global (as is standardly assumed), but local, in the sense that syntactic optimization procedures can affect only small portions of syntactic structure. Local optimization presupposes harmonic serialism (rather than harmonic parallelism), i.e., a derivational organization of grammar. In line with this, I set out to reconcile optimality theory with the minimalist program (see Chomsky (2000), Chomsky (2001)), a derivational approach in which phrase structure is created incrementally. I argue that local optimization is both conceptually attractive (because it significantly reduces complexity) and supported by empirical evidence. As a case study, I develop an analysis of a shape conservation phenomenon in German that involves repair-driven movement operations at the clause edge. I show that, other things being equal, local optimization succeeds where global optimization fails.

1. Background

Optimization can be parallel or serial, and it can be global or local. Optimization is parallel if it only applies once; it is serial if it applies more than once. Following Prince and Smolensky (1993), it is standardly assumed in optimality-theoretic phonology that optimization is parallel.[‡] In syntax, too, optimization is usually viewed as parallel.[§]

The issue of local vs. global optimization has so far received much less attention. An optimization is global if it affects the entire structure of a linguistic expression (e.g., word or sentence); it is local if it applies to a subpart of a linguistic expression. Most of the work in optimality theory relies on global optimization. This is particularly obvious in phonology, but it is also the case in syntax. However, local optimization in syntax is suggested as a possibility in Archangeli and Langendoen (1997, 214), and in a footnote in Ackema and Neeleman (1998, 478). Full-fledged analyses involving local optimization in syntax include Heck and Müller (2000a), Heck and Müller (2000b), Müller (2000), Fanselow and Čavar (2001), Heck (2001a), Fischer (2002), and Müller (2002).

Whereas a global approach can be either parallel or serial, a local approach must be serial, such that parts of sentences are successively subject to optimization. In what follows, I sketch a local optimization approach that incorporates main features of the minimalist program, whose incremental-derivational architecture makes it inherently serial.||

[‡] However, see McCarthy (2000), Rubach (2000), and the contributions in Hermans and van Oostendorp (2000) for (discussions of) serial optimization in phonology.

[§] See Grimshaw (1997), Pesetsky (1998), Legendre, Smolensky and Wilson (1998), Bresnan (2001), and most of the contributions in Barbosa et al. (1998), Legendre, Grimshaw and Vikner (1998), and Sells (2001). Exceptions that involve serial optimization are typically concerned with syntax/semantics interface phenomena (see, e.g., Heck (2001b) and Hendriks and de Hoop (2001)), and include various systems of bidirectional optimization (see Wilson (2001), Blutner (2000), Jäger and Blutner (2000), Aissen (2002), Lee (2001), Vogel (2002), Jäger (2002)). However, the number of optimization procedures required in these serial approaches is rather small (either 2 or 3).

|| Pesetsky (1998) and Broekhuis (2000) also combine assumptions of the minimalist program and optimality

2. Approach

Assume that syntactic structure is created incrementally from bottom to top as a result of derivational operations like Merge and Move that have access to the numeration (an array of items selected from the lexicon before the derivation starts). These operations belong to Gen, which also contains inviolable constraints, among them the Strict Cycle Condition (SCC) (Chomsky (1973), Chomsky (2001), Perlmutter and Soames (1979)) and the Phase Impenetrability Condition (PIC) (Chomsky (2000), Chomsky (2001)).

(1) *Strict Cycle Condition (SCC):*

Within the current XP α , an operation may not target a position that is included within another XP β dominated by α .

(2) *Phase Impenetrability Condition (PIC):*

The domain of a head X of a phase XP is not accessible to operations outside XP; only X and its specifier(s) are accessible to such operations.

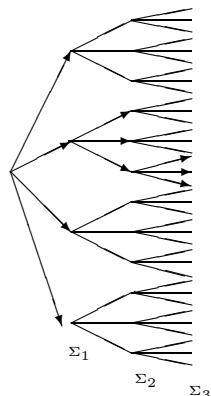
Focussing on Move operations here and in what follows, the SCC restricts the space in which the derivation can find a landing site for movement (i.e., locate the “probe”), whereas the PIC restricts the domain in which the derivation can find an item to move (i.e., a “goal”). The local domains are not completely identical: Every XP is a cyclic domain for the SCC, but only those XPs that qualify as phases (which for present purposes I assume to be CPs) are domains for the PIC.

Where, then, does optimization enter the picture? The idea is that certain derivational units act as local optimization domains Σ , and that Gen and H-Eval apply as many times as there are Σ s in the derivation. More specifically, suppose that on the basis of one and the same input, syntactic operations (Merge, Move, etc.) can apply in accordance with inviolable constraints (SCC, PIC, etc.) in different ways, yielding different outputs at stage Σ_i . These outputs are then subject to optimization along a set of ranked and violable constraints, and the optimal output is determined. Only an optimal output can show up in the input of subsequent derivational steps (together with items taken from the numeration and other optimal outputs), and the derivation proceeds in various Gen-compatible ways, producing different outputs at the next optimization domain Σ_{i+1} . At this point, optimization starts anew, yielding a winning candidate that acts as part of the new input, and so on, until all material of the numeration is used up, the derivation reaches an end, and the optimal root clause is determined. Importantly, all locally suboptimal outputs are disregarded in subsequent derivational steps. Therefore, local optimization significantly reduces complexity, compared with global optimization. This is shown schematically in figure 1.

Here, arrowed lines correspond to sequences of derivational steps that yield outputs which participate in local optimization. The other lines represent continuations of suboptimal outputs that give rise to many more outputs. These latter continuations and their associated outputs are simply not available in the local optimization approach adopted here; however, they must also be considered in a global approach, which cannot discriminate between arrowed and other lines. Consequently, a global approach is inherently more complex than a local approach. (It is worth emphasizing that this consequence arises in all global optimization approaches, independently of whether Gen is derivational or representational – in the latter case, the non-arrowed lines encode locally suboptimal subtrees.) Clearly, the degree to which local optimization and global optimization differ with respect to complexity depends on the choice of optimization domain: The smaller Σ is, the more local optimization pays off from

theory, but in a less far-reaching way that basically restricts H-Eval to PF-realization and relies on standard – parallel, global – optimization.

Figure 1: The size of candidate sets under local vs. global optimization



the point of view of complexity (Σ = root clause yields a global approach); on the other hand, an extremely small Σ brings with it the danger of leaving hardly any room for optimization (e.g., if Σ = derivational step). I assume that Σ = XP.

Another complexity-related issue arises if we assume serial optimization of growing subtrees: Does an output carry with it old (but, by definition, non-fatal) violation marks incurred by optimal outputs that are embedded in it, or are such violations invisible? It is not clear whether there is empirical evidence that would distinguish between the two options; but given the overall goal of reducing complexity, it seems preferable to assume that only those parts of an output are visible to H-Eval that are accessible in accordance with the PIC. Thus, only the structure from the present XP down to specifiers of minimally embedded CPs (if there are such) is subject to optimization at any given stage. More generally, we end up with the result that both the *number* and the *size* of competing syntactic outputs are considerably smaller than in systems that employ global optimization; taken together, these steps have the effect of bringing optimization in syntax closer to optimization in (non-phrasal) phonology and morphology, where such complexity issues are much less worrisome to begin with (essentially because words are smaller objects than sentences).

I take conceptual considerations like these to be suggestive; but eventually, the question of local vs. global optimization in syntax must be decided on the basis of empirical evidence. To this end, I provide an empirical argument for local optimization. The structure of the argument is as follows: (i) There is evidence for repair-driven movement at the edge of German clauses. (ii) Repair operations strongly suggest an underlying optimization procedure. (iii) The repair operation does not apply in all contexts in which the ranked constraints would seem to force it. (iv) The contexts in which it does not apply even though the constraints seem to demand application correspond to non-arrowed lines in figure 1 which are irrelevant in local optimization, but must be considered in global optimization.

The evidence I want to discuss involves a well-known asymmetry that shows up with *wh*-movement from embedded clauses in German.

3. Data

Two types of finite declarative clauses can be embedded under bridge verbs in German: (i) clauses headed by a complementizer *dass* ('that'); (ii) V/2 clauses with finite V in the C position and some XP in SpecC. Both types of complements as such appear to be transparent for *wh*-movement to SpecC. *Wh*-movement from a *dass* clause may go to a *dass* clause or to

a V/2 clause; see (3-ab).¶ In contrast, as shown in (3-cd), *wh*-movement from a V/2 clause may only end up in a V/2 clause again (see Tappe (1981), Haider (1984), Reis (1985)).

- (3) a. (Ich weiß nicht) [_{CP1} wen_i (dass) du meinst [_{CP2} t'_i dass sie t_i getroffen hat]]
 I know not whom that you think that she met has
 b. [_{CP1} Wen_i meinst du [_{CP2} t'_i dass sie t_i getroffen hat]] ?
 whom think you that she met has
 c. [_{CP1} Wen_i meinst du [_{CP2} t'_i hat sie t_i getroffen]] ?
 whom think you has she met
 d. *(Ich weiß nicht) [_{CP1} wen_i (dass) du meinst [_{CP2} t'_i hat sie t_i getroffen]]
 I know not whom that you think has she met

The same restriction holds when movement from SpecV/2 to Spec*dass* is followed by further *wh*-movement, or when the moved item is a topic or relative pronoun (the analysis below could be extended in obvious ways to cover topicalization and relativization). The data have proven remarkably robust over the years, and many attempts have been made to account for the asymmetry involved. First, it has been suggested that a V/2 clause acts as an island in (3-d), which then requires some extra assumption about (3-c), where islandhood seems to be voided (see Staudacher (1990), Sternefeld (1989), Reis (1996)). Second, it has been proposed that the asymmetry in (3) follows from directionality constraints on movement (see Müller (1989), Haider (1993)). Third, the data have been approached in terms of constraints against improper movement (see Haider (1984), Sternefeld (1992), Müller and Sternefeld (1993), Williams (2003)). However, all these approaches can be shown to involve construction-specific assumptions, and it seems fair to conclude that the problem in (3) has not yet received a satisfying solution.

4. Analysis

Suppose that movement is triggered by certain types of features on the probe that must be matched by appropriate features on the goal; following Sternefeld (2003), I refer to the features that trigger movement as [**F**] (i.e., “strong”) features, with matching [*F*] features on the goal. Two violable and ranked constraints play a role in this context: FC (*Feature Condition*) ensures that [**F**] on some lexical item X triggers movement to the edge of an XP (the edge of an XP comprises X and SpecX; see Chomsky (2000), Chomsky (2001)). LR (*Last Resort*) requires that movement results in feature matching.

- (4) a. *Feature Condition* (FC):
 An [**F**] feature on X requires an item bearing [*F*] at the edge of XP.
 b. *Last Resort* (LR):
 Movement requires matching of [*F*] and [**F**] at an edge.

Two further constraints of H-Eval are OP (*Operators at Clause Edges*; based on Grimshaw (1997)) and, crucially, SCE (*Shape Conservation for Clause Edges*). Shape Conservation has been suggested as a general constraint by Williams (2003).⁺ Versions of this constraint

¶ A complementizer of CP₁ must then be deleted in Standard German, but not in dialects and colloquial varieties. Following Pesetsky (1998), I assume that complementizer deletion is a PF phenomenon in languages like German and English, with a *that/dass* complementizer present in syntax proper.

⁺ Note that Williams (2003, 78-79) actually provides an account of the pattern in (3). However, Williams' analysis does in fact not rely on Shape Conservation; rather, it is an account in terms of improper movement that is very similar to the approach in Sternefeld (1992) – which in turn can be shown to be based on concepts that

are adopted within an optimality-theoretic approach in Müller (2001) (for co-argument NPs) and in Müller (2000) (for VPs). Thus, Shape Conservation can be viewed as a family of constraints, of which SCE is a member.

- (5) a. *Operators at Clause Edges* (OP):
An operator must be at the edge of a clause.
b. *Shape Conservation for Clause Edges* (SCE):
Clause edges have identical shapes.

SCE is a gradual constraint. Given the edge of a CP_α , SCE violations for CP_β are computed as follows: (i) Compare the n -th edge constituent of CP_α with the n -th edge constituent of CP_β and assign a * if the two items do not have an identical shape. (ii) For each edge constituent of one CP that does not correspond to an edge constituent of the other CP, assign a *.

Assume now a ranking $FC \gg OP$, $SCE \gg LR$. On this basis, let me first briefly address successive-cyclic *wh*-movement in general. Unbounded dependencies can be divided into three parts: a bottom, a middle, and a top (Gazdar et al. (1985)); see (6).

- (6) $\underbrace{[CP_1 wh_i C_{[*wh*]} \dots]}_{\text{top}} \underbrace{[CP_2 t_i'' C \dots]}_{\text{middle}} \underbrace{[CP_3 t_i' C \dots t_i \dots]}_{\text{bottom}}$

Movement at the top is triggered by FC, given a feature $[*wh*]$ on interrogative C and a matching feature $[wh]$ on a *wh*-phrase. In contrast, movement at the bottom and in the middle is not feature-driven (such intermediate movement steps are required theory-internally by the PIC; they are empirically supported by the existence of visible reflexes of successive cyclicity in the C domain in various languages). Movement that is not feature-driven violates LR; it qualifies as “repair-driven” in the terminology of Heck and Müller (2000a), i.e., it must be forced by a higher-ranked constraint. Movement at the bottom is triggered by OP, given the ranking $OP \gg LR$ (an “operator” in the sense of (5-a) is an XP that bears a feature like $[wh]$).

Finally, movement in the middle (a notorious problem in incremental-derivational approaches to syntax) is triggered by SCE, given the ranking $SCE \gg LR$.^{*} Here is why: Suppose that an $XP_{[wh]}-C$ shape has been created at the CP_α edge at the bottom. Then, SCE demands a replication of this shape at the next CP_β edge. As long as no higher-ranked constraint precludes this, the SCE thus triggers movement steps in the middle, in violation of LR.[‡] At the top, the demands imposed by SCE and FC converge. The question arises as to why SCE does not force *wh*-movement beyond a $[*wh*]$ target position (see Pullum (1979, 372)). This follows from the ranking $FC \gg SCE$: FC not only forces *wh*-movement to $SpecC_{[*wh*]}$; it also demands that the *wh*-phrase stays in this position.^{††}

Let me now turn to the specific situation in German. Suppose that $[*F*]$ features that can be on C include $[*xp*]$ (for movement of some XP to $SpecC$), $[*wh*]$ (for *wh*-movement), and $[*fin*]$ (for V/2-movement to C); these assumptions are virtually unavoidable in a feature-based approach to movement. Minimally, there must be two C elements in the lexicon for declarative clauses; these are rendered here as C_d and C_e : $C_d = [C \text{ dass}]$; C_d does not trigger

were first suggested in Williams (1974).

^{*} OP cannot force movement in the middle because it is satisfied once and for all when the *wh*-phrase has reached the first edge of a clause; see Fanselow and Ćavar (2001), who make use of this property of OP in their account of partial *wh*-movement constructions.

[‡] Isn't an $XP_{[wh]}-C$ shape of CP_α destroyed if $XP_{[wh]}$ moves to CP_β ? This issue does not arise if traces count for shape conservation. Alternatively, we can conceive of the shape of a CP edge as something that is fixed once and for all as soon as the CP has been optimized.

^{††} What about constructions in which NP-movement to subject position feeds *wh*-movement to $SpecC$? In these cases, there is no way to avoid a FC violation, and the decision then falls to independent constraints.

any movement via FC. $C_e = [C \emptyset_{[*xp*],[*fin*]}]$; C_e triggers V/2 and XP-movement to SpecC. Similarly, there are two C elements for interrogative clauses: $C_{dw} = [C \text{dass}_{[*wh*]}]$; C_{dw} attracts a *wh*-phrase via FC (and is PF-deleted in Standard German). $C_{ew} = [C \emptyset_{[*wh*],[*fin*]}]$; C_{ew} triggers *wh*-movement and V/2.

We can now derive the pattern in (3) on the basis of SCE. The two relevant local optimization procedures involve first the embedded CP_2 , and then the matrix CP_1 . SCE is always vacuously fulfilled in the first optimization procedure, and the optimal CP_2 will either be a *dass* clause or a V/2 clause, depending on the $[*F*]$ features of C. The competition in T_1 is based on an initial choice of C_d that is merged with the optimal TP created by the derivation so far; it produces an embedded *dass* clause as the optimal output, viz., O_2 ; only this output can then serve as an input for further operations. (Throughout, only the most relevant candidates are shown in tableaux.)

T_1 : ‘*dass*’ in CP_2 : (3-a), (3-b)

Input: $[C_d \text{dass}], [TP \text{sie wen getroffen hat}]$	FC	OP	SCE	LR
$O_1: [CP_2 [C \text{dass}] [TP \text{sie wen getroffen hat}]]$		*!		
$O_2: [CP_2 \text{wen}_i [C \text{dass}] [TP \text{sie } t_i \text{ getroffen hat}]]$				*
$O_3: [CP_2 \text{wen}_i [C \text{hat}_j (\text{dass})] [TP \text{sie } t_i \text{ getroffen } t_j]]$				**!

The derivation proceeds by optimizing the matrix VP and the matrix TP. Subsequent optimization of CP_1 may then lead to a *dass* clause or to a V/2 clause, depending on the nature of C as C_{dw} or C_{ew} ; see T_2 , T_3 . Consequently, (3-a) and (3-b) are both optimal. However, whereas optimal O_{22} in T_2 respects both FC and SCE (the clause edges have an identical $XP_{[wh]}-\text{dass}$ shape), optimal O_{24} in T_3 must violate SCE by applying V/2 in order to satisfy FC (for $[*fin*]$): *dass* is in C_2 , V/2 is in C_1 .[†]

T_2 : *Wh*-movement from ‘*dass*’ clauses into ‘*dass*’ clauses: (3-a)

Input: $[C_{dw} \text{dass}_{[*wh*]}], [TP \text{du meinst } [CP_2 \text{wen}_i [C \text{dass}] [TP \text{sie } t_i \text{ getroffen hat}]]]$	FC	OP	SCE	LR
$O_{21}: [CP_1 [C_{dw} \text{dass}] [TP \text{du meinst } [CP_2 \text{wen}_i [C \text{dass}] [TP \text{sie } t_i \text{ getroffen hat}]]]]]$	*!		**	
$O_{22}: [CP_1 \text{wen}_i [C_{dw} \text{dass}] [TP \text{du meinst } [CP_2 t'_i [C \text{dass}] [TP \text{sie } t_i \text{ getroffen hat}]]]]]$				

T_3 : *Wh*-movement from ‘*dass*’ clauses into V/2 clauses: (3-b)

Input: $[C_{ew} \emptyset_{[*wh*],[*fin*]}], [TP \text{du meinst } [CP_2 \text{wen}_i [C \text{dass}] [TP \text{sie } t_i \text{ getroffen hat}]]]$	FC	OP	SCE	LR
$O_{21}: [CP_1 [C_{ew} \emptyset] [TP \text{du meinst } [CP_2 \text{wen}_i [C \text{dass}] [TP \text{sie } t_i \text{ getroffen hat}]]]]]$	*!*		**	
$O_{22}: [CP_1 \text{wen}_i [C_{ew} \emptyset] [TP \text{du meinst } [CP_2 t'_i [C \text{dass}] [TP \text{sie } t_i \text{ getroffen hat}]]]]]$	*!			
$O_{23}: [CP_1 [C_{ew} \text{meinst}_j \emptyset] [TP \text{du } t_j [CP_2 \text{wen}_i [C \text{dass}] [TP \text{sie } t_i \text{ getroffen hat}]]]]]$	*!		**	
$O_{24}: [CP_1 \text{wen}_i [C_{ew} \text{meinst}_j \emptyset] [TP \text{du } t_j [CP_2 t'_i [C \text{dass}] [TP \text{sie } t_i \text{ getroffen hat}]]]]]$			*	

Consider now the case where the optimal embedded CP_2 is a V/2 clause, as in T_4 , which uses a different C_2 from T_1 , viz., C_e .

[†] Two remarks. First, the outputs are numbered O_{21} , O_{22} , ... so as to indicate that they are all descendants of O_2 in T_1 . Second, O_{22} in T_3 is here assumed to fully respect SCE; i.e., $[C \emptyset]$ and $[C \text{dass}]$ are taken to have identical shapes (as non-branching C items), in contrast to branching C items that result from V/2. However, this assumption is not crucial; a SCE violation in O_{22} in T_3 would not affect the outcome.

T_4 : V/2 in CP_2 : (3-c), (3-d)

Input: $[C_e \emptyset_{[*XP*],[*fin*]}], [TP \text{ sie wen getroffen hat }]$	FC	OP	SCE	LR
$O_1: [CP_2 [C_e \emptyset] [TP \text{ sie wen getroffen hat }]]$	*!*	*		
$O_2: [CP_2 \text{ wen}_i [C_e \emptyset] [TP \text{ sie } t_i \text{ getroffen hat }]]$	*!			
$O_3: [CP_2 [C \text{ hat}_j \emptyset] [TP \text{ sie wen}_i \text{ getroffen } t_j]]$	*!	*		
$\Rightarrow O_4: [CP_2 \text{ wen}_i [C_e \text{ hat}_j \emptyset] [TP \text{ sie } t_i \text{ getroffen } t_j]]$				
$O_5: [CP_2 \text{ sie}_k [C \text{ hat}_j \emptyset] [TP t_k \text{ wen}_i \text{ getroffen } t_j]]$		*!		

In this case, different choice of C_1 does *not* yield two different optimal outputs in CP_1 optimization. If C_1 has a $[*fin*]$ feature, the optimal CP_1 is also a V/2 clause because of FC, and SCE is respected; see T_5 . However, if C_1 does not have such a feature, V/2 will have to apply nonetheless – forced not by FC, but by SCE, in violation of LR; see O_{43} vs. O_{42} in T_6 . Here, we have an instance of repair-driven V/2 movement that gives rise to a neutralization effect. This derives the contrast between (3-c) and (3-d); the latter cannot be optimal.‡

T_5 : Wh-movement from V/2 clauses into V/2 clauses: (3-c)

Input: $[C_{ew} \emptyset_{[*wh*],[*fin*]}], [TP \text{ du meinst } [CP_2 \text{ wen}_i [C_e \text{ hat}_j \emptyset] [TP \text{ sie } t_i \text{ getroffen } t_j]]]$	FC	OP	SCE	LR
$O_{41}: [CP_1 [C_{ew} \emptyset] [TP \text{ du meinst } [CP_2 \text{ wen}_i [C_e \text{ hat}_j \emptyset] [TP \text{ sie } t_i \text{ getroffen } t_j]]]]$	*!*		**	
$O_{42}: [CP_1 \text{ wen}_i [C_{ew} \emptyset] [TP \text{ du meinst } [CP_2 t'_i [C_e \text{ hat}_j \emptyset] [TP \text{ sie } t_i \text{ getroffen } t_j]]]]$	*!		*	
$O_{43}: [CP_1 [C_{ew} \text{ meinst}_l \emptyset] [TP \text{ du } t_l [CP_2 \text{ wen}_i [C_e \text{ hat}_j \emptyset] [TP \text{ sie } t_i \text{ getroffen } t_j]]]]$	*!		**	
$\Rightarrow O_{44}: [CP_1 \text{ wen}_i [C_{ew} \text{ meinst}_l \emptyset] [TP \text{ du } t_l [CP_2 t'_i [C_e \text{ hat}_j \emptyset] [TP \text{ sie } t_i \text{ getroffen } t_j]]]]$				

T_6 : *Wh-movement from V/2 clauses into ‘dass’ clauses: (3-d)

Input: $[C_{dw} \text{ dass}_{[*wh*]}], [TP \text{ du meinst } [CP_2 \text{ wen}_i [C_e \text{ hat}_j \emptyset] [TP \text{ sie } t_i \text{ getroffen } t_j]]]$	FC	OP	SCE	LR
$O_{41}: [CP_1 [C_{dw} \text{ dass}] [TP \text{ du meinst } [CP_2 \text{ wen}_i [C_e \text{ hat}_j \emptyset] [TP \text{ sie } t_i \text{ getroffen } t_j]]]]$	*!		**	
$O_{42}: [CP_1 \text{ wen}_i [C_{dw} \text{ dass}] [TP \text{ du meinst } [CP_2 t'_i [C_e \text{ hat}_j \emptyset] [TP \text{ sie } t_i \text{ getroffen } t_j]]]]$			*!	
$\Rightarrow O_{43}: [CP_1 \text{ wen}_i [C_{dw} \text{ meinst}_l (\text{dass})] [TP \text{ du } t_l [CP_2 t'_i [C_e \text{ hat}_j \emptyset] [TP \text{ sie } t_i \text{ getroffen } t_j]]]]$				*

In a nutshell, then, the present analysis of the pattern in (3) is this: Given an optimal CP_2 , SCE demands that the edge of CP_1 has the same shape. This requirement can be met without problems in (3-a) and (3-c), where C_1 and C_2 are uniformly marked d (*dass*) or e (V/2). However, in (3-b) and (3-d), C_1 and C_2 differ with respect to d/e marking. This means that SCE can only be satisfied by violating some other constraint. In (3-b), this other constraint is

‡ C_e and C_{dw} in O_{43} of T_6 have identical shapes, as branching Cs. Note that the neutralization effect is not complete since O_{44} of T_5 has a \emptyset where O_{43} of T_6 has a *dass*. Hence, we would have to assume obligatory *dass* deletion at PF if O_{43} of T_6 could be (part of) a well-formed derivation – which, however, it can't be: O_{43} of T_6 can only be an intermediate optimal output (C_{dw} is always embedded and cannot be the head of a root clause); and there is a general prohibition against embedded *wh*-V/2 constructions in German (see Haider (1984)):

- (i) a. *Sie sagt $[CP \text{ wen}_i \text{ meinst du } t_i]$ b. Sie sagt $[CP \text{ wen}_i \text{ du } t_i \text{ meinst}]$
 she says who mean you she says who you mean

Accordingly, merging the optimal CP output of T_6 (or T_5 , for that matter) with V invariably results in ungrammaticality. Thus, independently of present considerations, there must be a high-ranked constraint \mathfrak{R} against merging V and a CP with $XP_{[wh]}$ -V/2 at its edge. Ineffability can then be derived in this context under a ranking $\mathfrak{R} \gg \text{EOC}$, where EOC is the *Empty Output Condition* that blocks the empty output \emptyset (the null parse). \emptyset is always present in competitions; its optimality signals a crash of the derivation (see Heck and Müller (2000a)).

FC; in (3-d), it is LR. Consequently, the ranking $FC \gg SCE \gg LR$ correctly predicts that SCE cannot stop feature-driven V/2 from applying in (3-b) (T_3), and that SCE forces repair-driven V/2 in (3-d) (T_6).

Needless to say, there are several further questions that will have to be addressed before the analysis can count as successful, and it will have to be extended in various ways.[§] Still, I would like to contend that the gist of the analysis in T_1 – T_6 can be maintained in a more comprehensive approach.

5. Argument

It remains to be shown that a global optimization approach would, *ceteris paribus*, fail in an analysis of the pattern in (3). This is straightforward: Under a global approach, we would wrongly expect SCE to require identity of the shape of clause edges much more generally, and could not account for the asymmetry observed in (3). In particular, (3-b) should be excluded in the same way as (3-d): CP_1 in O_{24} of T_3 violates SCE once; its predecessor CP_2 in O_2 of T_1 violates LR once. However, if the two CPs are optimized in parallel, the optimal output would combine CP_1 in O_{24} of T_3 and CP_2 in O_3 of T_1 (which is locally suboptimal because of a fatal LR violation due to locally unforced V/2). This would incur two violations of LR, but *no violation of SCE*; see T_7 , where the wrong winner (O_{34} , based on O_3) is marked ★, and well-formed O_{24} is blocked because of a fatal SCE violation. More generally, the global optimization approach predicts that an output at the right end of a non-arrowed line at a level like Σ_2 in figure 1 can be further used, and may ultimately lead to an output at a later level like Σ_3 that has a better constraint profile than the corresponding output at the right end of an arrowed line. This prediction is not borne out, though; hence, we have an argument against global optimization.

T_7 : Global optimization: *Wh-movement from ‘dass’ clauses into V/2 clauses: (3-b)

Input: [C_d dass], [TP sie wen getroffen hat] [C_{ew} \emptyset [$*wh*$], [$*fin*$]], [TP du meinst]	FC	OP	SCE	LR
O_{24} : [CP_1 wen _i [C_{ew} meinst _l \emptyset] [TP du t _l [CP_2 t' _i [C dass] [TP sie t _i getroffen hat]]]]			*!	*
★ O_{34} : [CP_1 wen _i [C_{ew} meinst _l \emptyset] [TP du t _l [CP_2 t' _i [C hat _j (dass)] [TP sie t _i getroffen t _j]]]]				**

6. Outlook

I have argued that a local approach to optimization in syntax is conceptually superior to a global approach because it reduces complexity; and I have shown that it also proves

§ To name just one relevant question: Why does SCE not force XP movement to SpecC in a matrix clause in the presence of ([$*xp*$]-driven) XP movement to SpecC in an embedded clause? In this context, there is no asymmetry between embedded *dass* clauses (as in (i-a)) and V/2 clauses (as in (i-b)); in particular, there is no repair-driven movement of both *er* and *sagte* to the edge of CP_1 in (i-b).

- (i) a. Ich denke [CP_1 er [C_e sagte] [CP_2 [C_d dass] sie schlafen möchte]]
I think he said that she sleep wants to
b. Ich denke [CP_1 [C_d dass] er sagte [CP_2 sie [C_e möchte] schlafen]]
I think that he said she wants to sleep

A simple solution would be to postulate a constraint \mathfrak{S} ($\mathfrak{S} \gg SCE$) that permits movement of a non-operator to the edge of C only if C is marked [$*xp*$]. On this view, movement theory is designed in such a way that only those items can move successive-cyclically that do in fact need to move in this manner, viz., operators.

empirically superior in the domain of successive-cyclic movement from *dass* vs. V/2 CPs in German, where it solves a recalcitrant problem via a simple Shape Conservation constraint.

The question arises of whether the local approach to optimization in syntax can be maintained in its strictest form (without adding limited look-ahead or backtracking capacity) in the light of other constructions that involve long-distance dependencies and thereby initially seem to support to a global approach. Phenomena that are relevant in this context include non-local reflexivization and resumptive pronoun strategies. Such non-local binding phenomena will have to be handled in a local approach by systematically decomposing non-local relations into a series of local feature passing operations (as proposed in Gazdar et al. (1985)), such that relevant information is accessible in each local optimization procedure. At the moment, I take it to be an open question whether this enterprise will ultimately be successful; however, preliminary results (see, e.g., Fischer (2003) on reflexives) suggest that such apparently non-local phenomena can indeed fruitfully be addressed in a local approach to optimization.

7. Bibliography

- Ackema, Peter and Ad Neeleman (1998): Optimal Questions, *Natural Language and Linguistic Theory* 16, 443–490.
- Aissen, Judith (2002): Bidirectional Optimization and the Problem of Recoverability in Head Marking Languages. Ms., University of California, Santa Cruz.
- Archangeli, Diana and Terence Langendoen (1997): Afterword. In: D. Archangeli and T. Langendoen, eds., *Optimality Theory. An Overview*. Blackwell, Oxford, pp. 200–215.
- Barbosa, Pilar, Danny Fox, Paul Hagstrom, Martha McGinnis and David Pesetsky, eds. (1998): *Is the Best Good Enough?*. MIT Press and MITWPL, Cambridge, Mass.
- Blutner, Reinhard (2000): Some Aspects of Optimality in Natural Language Interpretation, *Journal of Semantics* 17, 189–216.
- Bresnan, Joan (2001): The Emergence of the Unmarked Pronoun. In: G. Legendre, J. Grimshaw and S. Vikner, eds., *Optimality-Theoretic Syntax*. MIT Press, Cambridge, Mass., pp. 113–142.
- Broekhuis, Hans (2000): Against Feature Strength: The Case of Scandinavian Object Shift, *Natural Language and Linguistic Theory* 18, 673–721.
- Chomsky, Noam (1973): Conditions on Transformations. In: S. Anderson and P. Kiparsky, eds., *Festschrift for Morris Halle*. Academic Press, New York, pp. 232–286.
- Chomsky, Noam (2000): Minimalist Inquiries: The Framework. In: R. Martin, D. Michaels and J. Uriagereka, eds., *Step by Step*. MIT Press, Cambridge, Mass., pp. 89–155.
- Chomsky, Noam (2001): Derivation by Phase. In: M. Kenstowicz, ed., *Ken Hale. A Life in Language*. MIT Press, Cambridge, Mass., pp. 1–52.
- Fanselow, Gisbert and Damir Čavar (2001): Remarks on the Economy of Pronunciation. In: G. Müller and W. Sternefeld, eds., *Competition in Syntax*. Mouton de Gruyter, Berlin, pp. 107–150.
- Fischer, Silke (2002): Reanalyzing Reconstruction Effects: An Optimality-Theoretic Account of the Relation between Pronouns and R-Expressions. In: M. van Koppen, E. Thrift, E. J. van der Torre and M. Zimmerman, eds., *Proceedings of ConSole 9*. HIL, Leiden, pp. 69–81.
- Fischer, Silke (2003): Binding in a Derivational Approach. Ms., Universität Stuttgart. (Chapter 5 of forthcoming dissertation.).
- Gazdar, Gerald, Ewan Klein, Geoffrey Pullum and Ivan Sag (1985): *Generalized Phrase Structure Grammar*. Blackwell, Oxford.
- Grimshaw, Jane (1997): Projection, Heads, and Optimality, *Linguistic Inquiry* 28, 373–422.
- Haider, Hubert (1984): Topic, Focus, and V-Second, *Groninger Arbeiten zur Germanistischen Linguistik* 25, 72–120.
- Haider, Hubert (1993): ECP-Etuden: Anmerkungen zur Extraktion aus eingebetteten Verb-Zweit-Sätzen, *Linguistische Berichte* 145, 185–203.
- Heck, Fabian (2001a): Pied Piping Without Feature Percolation. Ms., Universität Stuttgart.
- Heck, Fabian (2001b): Quantifier Scopepe in German and Cyclic Optimization. In: G. Müller and W. Sternefeld, eds., *Competition in Syntax*. Mouton/de Gruyter, Berlin, pp. 175–209.
- Heck, Fabian and Gereon Müller (2000a): Repair-Driven Movement and the Local Optimization of Derivations. Ms., Universität Stuttgart and IDS Mannheim.
- Heck, Fabian and Gereon Müller (2000b): Successive Cyclicity, Long-Distance Superiority, and Local Optimization. In: R. Billerey and B. D. Lillehaugen, eds., *Proceedings of WCCFL*. Vol. 19, Cascadilla Press, Somerville, MA, pp. 218–231.

- Hendriks, Petra and Helen de Hoop (2001): Optimality Theoretic Semantics, *Linguistics and Philosophy* 24, 1–32.
- Hermans, Ben and Mark van Oostendorp, eds. (2000): *The Derivational Residue in Phonological Optimality Theory*. Benjamins, Amsterdam.
- Jäger, Gerhard (2002): Learning Constraint Sub-Hierarchies. The Bidirectional Gradual Learning Algorithm. Ms., ZAS Berlin/Universität Potsdam.
- Jäger, Gerhard and Reinhard Blutner (2000): Against Lexical Decomposition in Syntax. In: A. Wyner, ed., *Proceedings of IATL*. Vol. 15, University of Haifa, pp. 113–137.
- Lee, Hanjung (2001): Optimization in Argument Expression and Interpretation: A Unified Approach. PhD thesis, Stanford University.
- Legendre, Géraldine, Jane Grimshaw and Sten Vikner, eds. (1998): *Optimality-Theoretic Syntax*. MIT Press, Cambridge, Mass.
- Legendre, Géraldine, Paul Smolensky and Colin Wilson (1998): When is Less More? Faithfulness and Minimal Links in Wh-Chains. In: P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis and D. Pesetsky, eds., *Is the Best Good Enough?*. MIT Press and MITWPL, Cambridge, Mass., pp. 249–289.
- McCarthy, John (2000): Harmonic Serialism and Parallelism. In: M. Hirotani, A. Coetzee, N. Hall and J.-Y. Kim, eds., *Proceedings of NELS 30*. GLSA, Amherst, Mass., pp. 501–524.
- Müller, Gereon (1989): Barrieren und Inkorporation. Master's thesis, Universität Konstanz.
- Müller, Gereon (2000): Shape Conservation and Remnant Movement. In: M. Hirotani, A. Coetzee, N. Hall and J.-Y. Kim, eds., *Proceedings of NELS 30*. GLSA, Amherst, Mass., pp. 525–539.
- Müller, Gereon (2001): Order Preservation, Parallel Movement, and the Emergence of the Unmarked. In: G. Legendre, J. Grimshaw and S. Vikner, eds., *Optimality-Theoretic Syntax*. MIT Press, Cambridge, Mass., pp. 279–313.
- Müller, Gereon (2002): Harmonic Alignment and the Hierarchy of Pronouns in German. In: H. Simon and H. Wiese, eds., *Pronouns: Grammar and Representation*. Benjamins, Amsterdam, pp. 205–232.
- Müller, Gereon and Wolfgang Sternefeld (1993): Improper Movement and Unambiguous Binding, *Linguistic Inquiry* 24, 461–507.
- Perlmutter, David and Scott Soames (1979): *Syntactic Argumentation and the Structure of English*. The University of California Press, Berkeley.
- Pesetsky, David (1998): Some Optimality Principles of Sentence Pronunciation. In: P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis and D. Pesetsky, eds., *Is the Best Good Enough?*. MIT Press and MITWPL, Cambridge, Mass., pp. 337–383.
- Prince, Alan and Paul Smolensky (1993): *Optimality Theory. Constraint Interaction in Generative Grammar*. Book ms., Rutgers University.
- Pullum, Geoffrey (1979): *Rule Interaction and the Organization of a Grammar*. Garland, New York.
- Reis, Marga (1985): Satzeinleitende Strukturen im Deutschen. In: W. Abraham, ed., *Erklärende Syntax des Deutschen*. Narr, Tübingen, pp. 271–311.
- Reis, Marga (1996): Extractions from Verb-Second Clauses in German?. In: U. Lutz and J. Pafel, eds., *On Extraction and Extraposition in German*. Benjamins, Amsterdam, pp. 45–88.
- Rubach, Jerzy (2000): Glide and Glottal Stop Insertion in Slavic Languages: A DOT Analysis, *Linguistic Inquiry* 31, 271–317.
- Sells, Peter, ed. (2001): *Formal and Empirical Issues in Optimality Theoretic Syntax*. CSLI, Stanford.
- Staudacher, Peter (1990): Long Movement from Verb-Second-Complements in German. In: G. Grewendorf and W. Sternefeld, eds., *Scrambling and Barriers*. Benjamins, Amsterdam, pp. 319–339.
- Sternefeld, Wolfgang (1989): V-Movement, Extraction from V/2 Clauses, and the ECP, *Working Papers in Scandinavian Syntax* 44, 119–140.
- Sternefeld, Wolfgang (1992): Transformationstypologie und strukturelle Hierarchie. Ms., Universität Tübingen.
- Sternefeld, Wolfgang (2003): Syntax. Eine merkmalsbasierte Analyse. Book ms., Universität Tübingen.
- Tappe, Thilo (1981): Wer glaubst du hat recht?. In: M. Kohrt and J. Lernerz, eds., *Sprache: Formen und Strukturen*. Niemeyer, Tübingen, pp. 203–212.
- Vogel, Ralf (2002): Feedback Optimisation. Economy and Markedness in OT Syntax, Pt. I. Ms., Universität Potsdam.
- Williams, Edwin (1974): Rule Ordering in Syntax. PhD thesis, MIT, Cambridge, Mass.
- Williams, Edwin (2003): *Representation Theory*. MIT Press, Cambridge, Mass.
- Wilson, Colin (2001): Bidirectional Optimization and the Theory of Anaphora. In: G. Legendre, J. Grimshaw and S. Vikner, eds., *Optimality-Theoretic Syntax*. MIT Press, Cambridge, Mass., pp. 465–507.

The INPUT and Faithfulness in OT Syntax

Peter Sells

Stanford University

Abstract. I consider some of the claims that have been made for and against the nature of the INPUT in OT syntax as developed within the assumptions of the Minimalist Program, leading to suggestions for further specification of the architecture of this approach. Comparing with the role of faithfulness in the OT approach developed from Lexical-Functional Grammar, I argue that specific linguistic analyses crucially involve reference to faithfulness constraints (MAX and DEP in correspondence-based OT) which apply across different parts of the output structures, but do not need to refer to the INPUT. I conclude that while OT syntax does not need INPUTs per se, it does need faithfulness constraints.

1. Introduction

Much work in OT syntax has developed from roots in Government-Binding Theory (GB) or the Minimalist Program (MP), in particular following from the influential paper of Grimshaw (1997). I will refer to these as MP-OT approaches. Recently, the nature and role of the INPUT in MP-OT has come under scrutiny, especially in the survey of possibilities in Heck et al. (2002) (hereafter ‘Heck⁺’). They discuss difficulties in defining the INPUT in various MP-OT approaches, and suggest that a purely output-oriented alternative is possible, and preferable. In turn, this implies a rather radical difference in the architecture of the OT grammar between phonology and syntax, and Heck⁺ argue that this is not surprising due to important differences in faithfulness in the two domains. Primarily, they argue that syntax is almost fully faithful, and that the effects of faithfulness constraints on classical INPUT-OUTPUT relations can be recoded as enrichments solely in the output structures.

In this paper I will consider some of the claims that have been made for and against the nature of the INPUT in MP-OT, with a goal to suggesting some elaborations that can be made in specifying the MP-OT architecture.[†] I will also discuss the somewhat different approach to OT syntax that has developed out of Lexical-Functional Grammar (LFG), in the system that I will refer to as LFG-OT (see Bresnan (2000)). Although LFG-OT has a well-defined notion of INPUT, Kuhn (2001b) has shown that it does not constitute a crucial part of the system. However, specific analyses may involve reference to faithfulness constraints (MAX and DEP in correspondence-based OT) which apply across different parts of the output structures. I will argue that the results from such analyses cannot easily be recoded in the style alluded to by Heck⁺. The conclusion, then, will be that while OT syntax does not need INPUTs, it does need faithfulness constraints. This contrasts with the stated goal of Heck⁺ (2002, 363) which is “to reconstruct different theories without making use of the notion of faithfulness”.

2. Non-Convergent Derivations

It has been the practice in MP-OT to take constraints which must necessarily be satisfied for convergence in standard MP work and reinterpret them as violable. A problem with this MP-OT conception of constraint violation is that it leads to ill-formed representations

[†] I am grateful to Gereon Müller for useful comments on the first draft of this paper.

which cannot be part of convergent derivations. While this has no unwanted consequences for some constraints, in general it leads to analyses that involve improper representations and/or derivations; for example, consider the constraints in (1):

- (1) a. STAY: Do not move (or: **t*) (from Grimshaw (1997)).
- b. OP-SPEC: A *wh*-phrase must be in specifier position (from Grimshaw (1997)).
- c. CASE: A DP must check Case (cf. Grimshaw (1997), Speas (2001), Vikner (2001)).

Consider the effects of STAY. In a grammar where STAY is ranked very high, there will be little movement, if any. However, all MP approaches rely on movement of every piece of structure from its position of initial Merge, for the purposes of syntactic feature-checking.[‡] For example, a *wh*-phrase should move to check its *wh*-feature, and to put itself in an acceptable scope position at LF (to respect OP-SPEC). And a DP should move to be the specifier of some functional head – if it does not move to check its Case (thereby respecting STAY), the result is ill-formed at least at LF and also possibly at PF; in other words, it is part of a non-convergent derivation. Yet STAY could prevent these movements from happening (e.g., the simplest analysis of a *wh*-in-situ language is STAY \gg OP-SPEC).

If non-convergence is allowed, it is perhaps possible that an OT grammar could be defined, with a suitable INPUT. For example we might imagine that for some INPUT LF_i there are candidate derivations D_1, D_2, \dots etc., and that even if D_i does not converge, it is ‘far enough’ along its derivational path to LF_i that it might be the best candidate (though unfaithful). However, it seems problematic in the general case to specify which particular LF(s) a non-convergent derivation will yield. So, either the derivation must be allowed to continue beyond the points of Spell-Out to avoid non-convergence, in which case the role of the OT constraints is considerably reduced, or the derivational part of MP-OT must be reformulated; Broekhuis and Dekkers (2000, 418–421) discuss these issues in the context of their own OT approach which gives only a limited role to the OT evaluation itself (see also Broekhuis (2000)). In the following section, I suggest that MP-OT should adopt a different but already-existing model of the the derivation that avoids the non-convergence problem.

3. Revisions to the mechanisms of MP-OT

3.1. INPUTS or Enriched OUTPUTs?

Heck⁺ survey various approaches to the INPUT in MP-OT. If the OT analysis starts with just a numeration, it is not possible to define the candidate set properly, for the same numeration may lead to different LFs, and different numerations may lead to the same LF. Intuitively, though, some notion of LF-comparability seems to be what is needed (Heck⁺, fn. 16). So, following e.g., Legendre et al. (1998), it seems that the INPUT must be a predicate-argument structure with all relevant (semantic) features of the arguments specified, plus an indication of target scope for any potentially scopal elements, and probably a similar indication of Information Structure status (e.g., Topic and Focus). As an INPUT structure is quite articulated – it would have to be in order to be compared to an LF for the determination of faithfulness – the question Heck⁺ raise is this: where does the structured INPUT come from, or, what principles govern its well-formedness?

Heck⁺ argue that this question can simply be avoided, by moving to output-oriented OT. The candidate set is chosen in terms of all derivations which yield the same interpretation. Heck⁺’s strategy for OT syntax is to enrich the OUTPUT structures so that they encode

[‡] This may not be true of the AGREE version of MP of Chomsky (2000).

enough information that the INPUT can be eliminated, offering this view: “Syntactic output representations are richly structured objects that can provide all the information that faithfulness constraints locate in the input” (p. 371). For example, a derivation in which some scopal element stays in-situ will have the property that that scopal element is not positionally faithful to its intended scope. If so, how is its ‘intended scope’ determined, when there is no separate structured INPUT? – Heck⁺ propose to augment the syntactic structures with scope indicators (pp. 368–9).

However, there are three problems with this approach. The first is that it involves giving up the Inclusiveness Condition of Chomsky (1995, 228) (see (3)b), for scope markers are presumably not part of the initial numeration. The second problem is that a derivation with scope indicators in it is not the same as an LF that marks scope; in other words, there has to be some implicit external standard by which different derivations are judged to have the same LF (this assumes that the problem of non-convergent derivations is solved). But to avoid this ‘comparability’ issue, it would seem that every (legal) derivation should yield at least one LF, rather than being a derivation which may or may not yield an LF and may or may not have scope markers in it. The third problem is a technical one: unless the derivation continues to the point where every scopal element is actually assigned the scope that its scope marker indicates, there is no mechanism to determine whether an unmoved scopal element can legitimately take the scope of the position where its scope marker actually is. In other words, given that scope markers are not parts of normal syntactic derivations, they have to be inserted into structures (somehow) and ‘eliminated’ by scopal elements taking their scope through legal movements. Again, this suggests that every derivation continues to LF, regardless of the PF-position of the overt elements.

3.2. *Single Output Syntax*

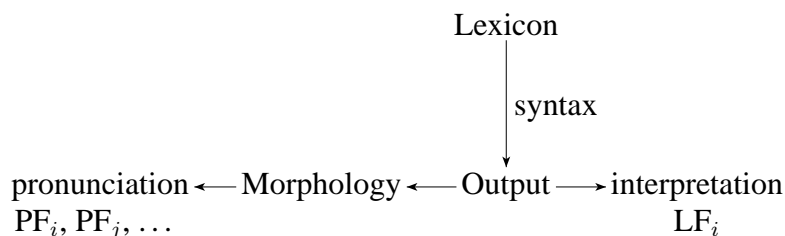
At a very general level, analyses that have been given in MP-OT allow constraint violations to give structures which could be described by the statements in (2), but the MP properties in (3) have generally been considered to be non-negotiable.

- (2) a. X is in a marked position (the ‘wrong’ position for feature checking).
- b. X has made an illegal move (X is in the wrong position with respect to where it started).
- (3) Inviolable MP Principles:
 - a. Start with a numeration.
 - b. Respect the Inclusiveness Condition. (No new objects are added in the course of computation.)
 - c. Structure is built by Merge and Move.
 - d. X'-theory (Bare Phrase Structure) holds uniformly.

What is interesting about these is that (2) essentially concerns PF-properties – for whatever reason, X is in the wrong position – while (3) concerns the dynamics of the computational system itself. This suggests the adoption of a particular kind of Minimalist syntax dubbed ‘Single Output Syntax’ by Bobaljik (2002) (see also Groat and O’Neil (1996)), which I now briefly describe.

The Copy Theory of movement of Chomsky (1995) leaves a copy of each moving element in situ; movement creates a chain of copies, and the highest copy is pronounced. Some MP researchers have proposed Single Output Syntax, in which a derivation proceeds to LF, with different Spell-Outs creating different PFs, depending on which copy in a chain is

Figure 1. Single Output Syntax



pronounced. For example, a DP that needs to check Case always moves from its base position to the position where its Case is checked, but it may undergo Spell-Out in either the higher or lower position, corresponding to overt or covert movement in the system of Chomsky (1995). The overall system generates pairs $\langle LF_i, PF_i \rangle$, $\langle LF_i, PF_j \rangle$, etc.; each is convergent to LF, and therefore the set of such pairs can be considered candidates if the system is extended to one of OT competition. The architecture of the system is shown in Figure 1. §

4. Faithfulness in LFG-OT

LFG-OT (see Bresnan (2000), Kuhn (2001b), Sells (2001)) has a well-defined INPUT and OUTPUT. The LFG framework is based on a correspondence theory between the overt structure (c-structure) and an abstract structure representing language-invariant information (f-structure), each of which is generated by independent principles and subject to specific general well-formedness constraints (see Kaplan and Bresnan (1982)). In LFG-OT pairs of $\langle \text{c-structure}, \text{f-structure} \rangle$ define the candidates in the OUTPUT, and the INPUT is defined as a skeletal f-structure. Simplifying a little, Kuhn proposes a system in which GEN relates an INPUT f-structure to the candidates in this way: each candidate is a c-structure/f-structure pair as just described such that the INPUT subsumes the f-structure part of the candidate, as shown in Figure 2. GEN may further specify the INPUT information but cannot eliminate any of it. The architecture is illustrated with the example ‘Anna has read novels’ in Swedish, *Anna har läst romaner*, in anticipation of the discussion in section 5.

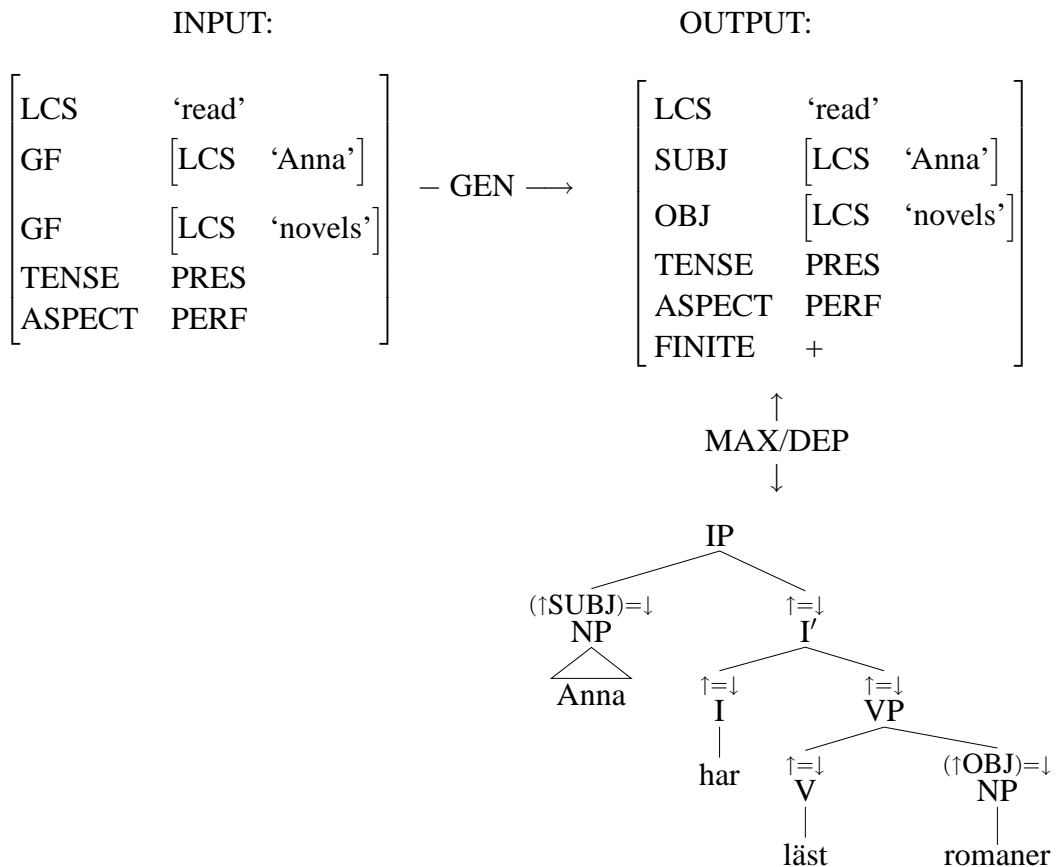
However, faithfulness constraints only hold between the two parts of the OUTPUT pair:

- (4) Faithfulness violations in LFG-OT
 - a. MAX: F-structure information does not have corresponding c-structure expression.
 - b. DEP: Lexically-specified information on a c-structure (terminal) node is not present in the f-structure.

Illustrating with simple examples, the Italian *canta* ‘(She) sings.’ is a MAX violation because the third-person pronoun information about the subject has no correspondent in the c-structure, while *do* in *I do not sing* is a DEP violation because the LCS information associated with the verb *do* does not appear in the f-structure of the example (for its LCS value is ‘sing’). This account is available in LFG-OT because f-structures and c-structures are defined independently of each other, and the syntactic theory establishes correspondences between them. So, one can look at the information in an f-structure as described by the MAX

§ Due to Paul Hagstrom, from <http://www.bu.edu/linguistics/UG/course/lx523-s01/handouts/SyntaxII.6.OS.Bobaljik.pdf>.

Figure 2. OT-LFG



constraint and see if there is a c-structure correspondent of it. In the other direction, so to speak, one can look at a c-structure node and see if all of the information associated with it is matched by or projected to information in f-structure (for DEP). Kuhn (2001a, 2001b) refers to this as the ‘lexicalist view of faithfulness’.

5. Faithfulness in MP-OT

The interpretation given to MAX and DEP constraints in LFG-OT can be carried over to MP-OT with certain revisions to the underlying architecture. Before getting to that, it is necessary to re-evaluate the arguments in Heck⁺ that syntax is largely faithful and that what residue of unfaithfulness there is can be accommodated into enriched OUTPUTs. Heck⁺ propose to deal with the two basic cases of faithfulness violations as in (5):

- (5) a. MAX: the structure involves empty categories rather than overt categories (hence, structure in the candidate is only apparently ‘missing’).
b. DEP: avoid expletive elements.

MAX constraints concern phenomena traditionally handled by ellipsis or deletion. Heck⁺ offer one strategy of assimilating all such cases to an alternation between overt and covert categories. The classic case of Pro-Drop mentioned above illustrates: in certain syntactic contexts, *pro* can be used instead of an overt pronoun. A different strategy for such missing arguments would be to check that the arguments in the verb's θ -grid are all realized in syntax:

- (6) All arguments of the verb are realized in syntax. (Heck⁺ (33)c, p. 365)

This would be a ‘DEP’ constraint (see also (8) below). Such an approach compares the information in the individual components of the derivation with how that information is used in the actual derivation, and hence has the ‘lexicalist view’ on the faithfulness just mentioned, a point of convergence in the MP-OT and LFG-OT approaches. Unfortunately, though, it seems to lead back to the non-convergence problem discussed in section 2: if an argument of the verb is unrealized in syntax, the LF will violate Full Interpretation or other well-formedness constraints (see Chomsky (1995, 220)). Additionally, such an LF will incorrectly fail to compete with an identical LF that does have all arguments realized.

5.1. Empirical Arguments for Faithfulness

A serious problem for the Heck⁺ proposal comes from the DEP constraints. There are expletive uses of pronouns like *it* and *there*, and there are expletive uses of verbs. Following the analysis of *do* in Grimshaw (1997), it seems straightforward to show that any auxiliary use of a contentful verb is an expletive use in the same sense. For example, *haben* is used in German as one expression of the simple PAST, and *werden* is used to mark the FUTURE:

- | | | | |
|--------|---|----|--|
| (7) a. | Er hat getanzt.
he have.Pres dance.Part
‘He danced.’ | b. | Er wird tanzen.
he become.Pres dance.Inf
‘He will dance.’ |
|--------|---|----|--|

These are periphrastic expressions of simple tenses. Ignoring the complication of the morphological forms (both are morphologically present tense sentences but one means PAST and one means FUTURE; see e.g., Ackerman and Webelhuth (1998)), the problem is that the LCSs of *haben* and *werden* are not part of the semantic interpretation of the examples. It would be possible to mark each verb as ‘expletive’, but then it would be a mystery (just as it is with the pronominal expletives) that these auxiliary uses of verbs have exactly the same range of surface forms as they do in their ‘main’ verb uses. In other words, at some level, main verb *haben* and auxiliary *haben* are both *the same verb*. This is precisely what the Grimshaw-approach captures, and it is formalized as a DEP constraint by Kuhn. Referring back to Figure 2, the Swedish verb *har* is present in c-structure but its LCS does not appear in the f-structure, which is a DEP violation.

How would this work in MP-OT? It would mean that *haben* is in the numeration (along with *gehen*) but that part of its lexical information, its LCS, is not used in the derivation. Alternatively, in a modified Distributed Morphology approach (Halle and Marantz (1993), Wiklund (2001)), the syntax might just manipulate syntactic features, but then the Vocabulary Insertion for German past and future tenses would involve insertion of just the phonological form of a contentful verb. (In Figure 1, this would mean that the Morphology component accesses the Lexicon.) In fact, Heck⁺ effectively adopt a faithfulness-based proposal, as they assume access to lexical information – information in a lexical element drawn from the numeration – proposing (8) (their (30)c):

- (8) Lexical Conceptual Structure is parsed.

Nevertheless they claim that this is not a faithfulness constraint (even though it uses the term ‘parse’). The reason is that they assume that the auxiliary verb is picked up during the course of the derivation, and so is not part of the INPUT in the usual sense. This seems like a terminological quibble: certainly a verb is chosen but not all of its information is used.¶

¶ As Kuhn (2001b) emphasizes, there are two senses of INPUT in derivational OT syntax: the INPUT as the starting point or target of the derivation, and the INPUT as the set of lexical resources (the numeration) used in a derivation. Everyone agrees that only the latter is viable, and papers such as Heck⁺’s and this one are investigating that sense of INPUT.

A stronger argument for faithfulness can be made, on the basis of *ha*-deletion in the mainland Scandinavian languages (see den Besten (1983), Platzack (1986) and Holmberg (1986), among others). Illustrating with Swedish data, the generalization is that *ha* is omissible in construction with a Supine form of the main verb just in case *ha* is not needed to be the finite verb in second position in a V2 clause. Consequently, *ha* is omissible in (9)a but not (9)b, and even though *har* is in second position in the embedded clause in (10), it is omissible, as this is not a V2 clause.

- (9) a. Han måste (ha) varit sjuk.
 he must.Past (have.Base) be.Sup sick
- b. Han *(har) varit sjuk.
 he *(have.Pres) be.Sup sick
- (10) Jag tror [att han (har) varit sjuk].
 I think.Pres [that he (have.Pres) be.Sup sick]

Finally, even though (11) is a V2 clause, the ‘finite’ function is exceptionally fulfilled by the adverb *kanske* (see e.g., Platzack (1986)), and *har* is again optional.

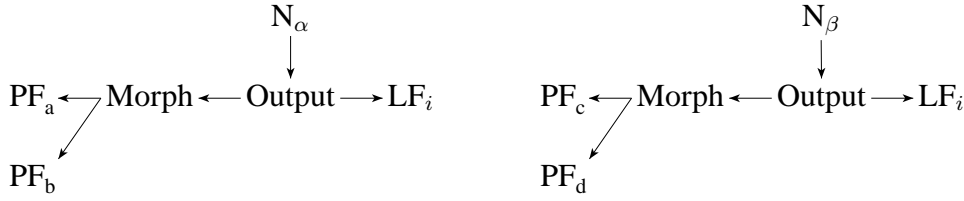
- (11) Allan kanske (har) redan ätit frukost.
 Allan maybe (have.Pres) already eat.Sup breakfast

In Sells (2003) I argued that the pressure to delete *ha* comes from the fact that using it is a DEP violation, as described above, and further that relating *ha*-omission to a DEP violation necessarily limits the phenomenon to auxiliary uses of verbs, correctly (cf. den Besten (1983) on *sein*-omission in German). And in some clauses, when *ha* is absent, there is no direct exponent of finiteness at all (e.g., the embedded clause in (10) without *ha*). I argued that this fact motivates a constraint ranking DEP(LCS) \gg MAX(FIN) – in other words, it is better to avoid the DEP violation than to express the finiteness of the clause, which would otherwise require the finite form *har*. Looking back to Figure 2, if *har* were absent, there would be no part of the c-structure expressing the [FINITE +] information in the f-structure, which would be a MAX violation with respect to that particular attribute.

Under the Heck⁺ conception of MAX constraints, the alternation would not be between *ha* and its absence, but between *ha* and a zero-alternate, \emptyset_{HA} (just like a pronoun vs. *pro*). Potentially, there would have to be several such zero verbs (e.g., the German data above would motivate \emptyset_{HABEN} and \emptyset_{WERDEN}). In the Swedish data in question, \emptyset_{HA} would be present in *ha*-less examples. However, \emptyset_{HA} would not save a DEP violation, for it has an LCS just like *ha* does. Under this approach, there would be no explanation of why it is only auxiliaries (i.e., main verbs used without their semantic content) in periphrastic expressions that are the omissible verbs. To preserve the faithfulness-based account, it is necessary to treat at least this case of absence as literal absence – there is no verb at all, not even a zero verb.

Now, Julien (2000) argues that there is a ‘recoverability’ constraint on deletion of finite *ha*: it is only omissible when there is some other indication that the clause is finite. For example, with *har* absent in (10), the fact that the subject has nominative Case still shows that the embedded clause is finite. Julien observes that *ha*-deletion is not possible in embedded clauses where *ha* itself would be the only overt reflex of finiteness. Sells (2003) interprets Julien’s insight as a DEP(FIN) constraint, to the effect that evidence of finiteness in the overt string must be matched by finiteness in the abstract structure. In the bidirectional LFG-OT account in Sells (2003), a c-structure with *han* in it must be associated with an f-structure with [FINITE +] in it, otherwise there is a DEP violation.

Figure 3. MP-OT Derivations



For illustration, assume two numerations, N_α and N_β , both of which participate in derivations that lead to the same LF, and both of which lead to two separate PF outputs, giving 4 $\langle LF_i, PF \rangle$ pairs. These will be candidates in an OT evaluation.

5.2. Faithfulness in MP-OT

I propose that we modify the MP-OT approach of Heck⁺ to allow for complete absence, to give the right results for DEP(LCS). The question is now how to handle the faithfulness constraints just described for finiteness. This grammatical property will be associated with the Fin functional head, proposed by Rizzi (1997) – a clause is finite if it has a certain specification of Fin(P). The Swedish examples show that there are two ways in which Fin can participate in the structure: first, Fin has certain properties within the derivation, e.g., it may be involved in Case checking, or it may be selected by certain complementizers, which correspond to Julien’s ‘recoverability’ evidence; and second, the Fin head must find an overt expression (cf. the ‘Fin*’ proposal of Roberts and Roussou (2002)). In terms of OT constraints, DEP(FIN) concerns the first property and MAX(FIN) concerns the second. MAX(FIN) means that finiteness should be expressed, in other words, that Fin is really Fin* (using the Roberts/Roussou notation). We have seen that *ha*-deletion happens when MAX(FIN) is outranked by DEP(LCS), but Julien observed that in those cases finiteness must still be recovered. This is the effect of DEP(FIN), which looks for some syntactic evidence of Fin in the output (even if the Fin head was not overtly expressed).

Specifically, the derivation proceeds from a numeration N but may not use all the information in N , so there are many derivations from a single N (in turn, many of which will lead to non-convergence and/or different LFs). However, the derivation need not add information to the elements drawn from N , respecting the Inclusiveness Condition. The relation between N and the derivation (see Figure 3) is moderated by faithfulness constraints, given informally in (12):

- (12) Faithfulness in MP-OT
- a. DEP(F): a feature F of an item in N has a role in the derivation.
 - b. MAX(F): a feature F in an item in N is expressed at PF.

The intent behind these has been explained above. DEP requires that each property of each item in N be ‘used’ – it must participate in selection, checking, or interpretation, for example. And those features which Roberts and Roussou (2002) notate with * to signify the need for overt expression can be dealt with by MAX constraints. Looking at Figure 3, we can imagine that N_α and N_β are different numerations, of which only a subset of the information is actually used in the derivations to LF_i . Hence the candidates will have different DEP constraint profiles, and, as they have different PFs, by assumption, they may have different profiles for MAX constraints. They will also have different profiles for the structural markedness constraints which dictate what the PFs will be. In other words, the pairs that form the candidate set can be evaluated against each other with respect to markedness and faithfulness.

These considerations on the relation between lexical resources and the derivation also seem relevant for the serial local optimization approach to MP-OT presented in Müller (2003). In this approach each local derivation of structure is the input for the next derivational cycle. Nevertheless lexical resources have to be taken from the numeration and used in the derivation, and auxiliary verbs, for example, seem to have the same status and properties of unfaithfulness that I have discussed above.

6. Conclusion

In summary, I have argued that both MP-OT and LFG-OT produce syntactic analyses which relate an overt form (PF or c-structure) to a largely language-invariant abstract structure (LF or f-structure). Neither MP-OT nor LFG-OT need an INPUT as classically defined in OT, for the candidate set can be defined directly in terms of equivalence at the LF/f-structure level. However, faithfulness cannot be dispensed with, even though the INPUT can. Insightful analyses of expletives and of recoverability require MAX and DEP constraints, in either approach. Taking a lead from some of the LFG-OT work, I have suggested that a good model for the MP-OT derivational system is that of Bobaljik (2002) (among others), in which the derivation ‘continues’ to LF no matter where individual items are subject to Spell-Out, as shown schematically in Figure 1.

In terms of the bigger picture, the considerations here agree with the claim in Heck⁺ that the system of OT syntax differs from the system of OT phonology, but also provide a strong case that a ‘lexicalist view’ of faithfulness in syntax should not be dispensed with.

References

- Ackerman, Farrell, and Gert Webelhuth. 1998. *A Theory of Predicates*. Stanford, CSLI Publications.
- Bobaljik, Jonathan D. 2002. A-chains at the PF interface: Copies and ‘covert’ movement. *Natural Language and Linguistic Theory* 20, 197–267.
- Bresnan, Joan. 2000. Optimal syntax. In Joost Dekkers, Frank van der Leeuw, and Jeroen van de Weijer (eds.), *Optimality Theory: Phonology, Syntax and Acquisition*. Oxford, Oxford University Press, 334–385.
- Broekhuis, Hans. 2000. Against feature strength: The case of Scandinavian Object Shift. *Natural Language and Linguistic Theory* 18, 673–721.
- Broekhuis, Hans, and Joost Dekkers. 2000. The Minimalist Program and Optimality Theory: Derivations and evaluations. In Joost Dekkers, Frank van der Leeuw, and Jeroen van de Weijer (eds.), *Optimality Theory: Phonology, Syntax and Acquisition*. Oxford, Oxford University Press, 386–422.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MIT Press.
- Chomsky, Noam. 2000. Minimalist inquiries: the framework. In Roger Martin, David Michaels, and Juan Uriagereka (eds.), *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*. Cambridge, MIT Press, 89–155.
- den Besten, Hans. 1983. On the interaction of root transformations and lexical deletive rules. In Werner Abraham (ed.), *On The Formal Syntax Of The Westgermania*. Amsterdam/Philadelphia, John Benjamins, 47–131.
- Grimshaw, Jane. 1997. Projection, heads, and optimality. *Linguistic Inquiry* 28, 373–422.
- Groat, Erich, and John O’Neil. 1996. Spell-out at the LF interface. In W. Abraham, S. Epstein, H. Thráinsson, and C. J.-W. Zwart (eds.), *Minimal Ideas*. Amsterdam/Philadelphia, John Benjamins, 113–139.

- Halle, Morris, and Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In K. Hale and S. J. Keyser (eds.), *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*. Cambridge, MIT Press, 111–176.
- Heck, Fabian, Gereon Müller, Ralf Vogel, Silke Fischer, Sten Vikner, and Tanya Schmid. 2002. On the nature of the input in Optimality Theory. *The Linguistic Review* 19, 345–376.
- Holmberg, Anders. 1986. *Word Order and Syntactic Features in the Scandinavian Languages and English*. Stockholm, University of Stockholm, Department of Linguistics.
- Julien, Marit. 2000. Optional *ha* in Swedish and Norwegian. *Working Papers in Scandinavian Syntax* 66, 33–74.
- Kaplan, Ronald M., and Joan Bresnan. 1982. Lexical-Functional Grammar: A formal system for grammatical representation. In Joan Bresnan (ed.), *The Mental Representation of Grammatical Relations*. Cambridge, MIT Press, 173–281.
- Kuhn, Jonas. 2001a. Generation and parsing in Optimality Theoretic syntax: Issues in the formalization of OT-LFG. In Peter Sells (ed.), *Formal and Empirical Issues in Optimality Theoretic Syntax*. Stanford, CSLI Publications, 313–366.
- Kuhn, Jonas. 2001b. *Formal and Computational Aspects of Optimality-Theoretic Syntax*. Doctoral dissertation, Institut für maschinelle Sprachverarbeitung, Universität Stuttgart.
- Legendre, Géraldine, Paul Smolensky, and Colin Wilson. 1998. When is less more? Faithfulness and minimal links in *wh*-chains. In Pilar Barbosa et al. (ed.), *Is the Best Good Enough? Optimality and Competition in Syntax*. Cambridge, MIT Press, 249–289.
- Müller, Gereon. 2003. Local vs. global optimization in syntax: A case study. This volume.
- Platzack, Christer. 1986. COMP, INFL, and Germanic word order. In Lars Hellan and Kirsti Koch Christensen (eds.), *Topics in Scandinavian Syntax*. Dordrecht, Reidel, 185–234.
- Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.), *Elements of Grammar*. Dordrecht, Kluwer Academic Publishing, 281–337.
- Roberts, Ian, and Anna Roussou. 2002. The Extended Projection Principle as a condition on the Tense Dependency. In Peter Svenonius (ed.), *Subjects, Expletives, and the EPP*. New York, Oxford University Press, 125–155.
- Sells, Peter (ed.). 2001. *Formal and Empirical Issues in Optimality Theoretic Syntax*. Stanford, CSLI Publications.
- Sells, Peter. 2003. Morphological and constructional expression and recoverability of verbal features. Ms. Stanford University. To appear in C. Orhan Orgun and Peter Sells (eds.) *Morphology and the Web of Grammar: Essays in Memory of Steven G. Lapointe*. Stanford, CSLI Publications.
- Speas, Margaret. 2001. Constraints on null pronouns. In Géraldine Legendre, Jane Grimshaw, and Sten Vikner (eds.), *Optimality-Theoretic Syntax*. Cambridge, MIT Press, 393–425.
- Vikner, Sten. 2001. V^0 -to- I^0 movement and *do*-insertion in Optimality Theory. In Géraldine Legendre, Jane Grimshaw, and Sten Vikner (eds.), *Optimality-Theoretic Syntax*. Cambridge, MIT Press, 427–464.
- Wiklund, Anna-Lena. 2001. Dressing up for Vocabulary Insertion: the Parasitic Supine. *Natural Language and Linguistic Theory* 19, 199–228.

Participant reduction and two-level markedness

Jochen Trommer

Institute of Cognitive Science, University of Osnabrück

1. Overview

Distributed Optimality (Trommer 2001), unlike standard Correspondence Theory (McCarthy and Prince 1994), claims that markedness constraints can refer to input *and* output representations. In this paper, I discuss the phenomenon that number features in transitive agreement with two speech act participants (SAPs, 1st and 2nd person arguments) are neutralized ("Participant Reduction") and argue that this effect is due to the constraint Participant Uniqueness (P.U.). Based on data from Colloquial Ainu (Shibatani 1990), I show that P.U. favors unfaithful candidates with reference to input features and provides evidence for two-level markedness constraints at the morphology-syntax interface.

2. Participant reduction in colloquial Ainu

In Colloquial Ainu (Shibatani 1990, p 29), subject and object agreement in transitive forms is marked transparently by prefixes, where subject precede object prefixes:

- (1)
- | | | |
|----|--------------------|---------------------|
| a. | <i>eci-un-kore</i> | 'you (pl.) give us' |
| | 2-O1p-give | |
| b. | <i>e-en-kore</i> | 'you (sg.) give me' |
| | 2sg-O1s-give | |

However, in all combinations, where the subject is 1st and the object 2nd person, only the 2nd person marker *eci-* appears (2). The left column contains the compositional forms that would be expected (*ku-*, S1sg; *ci-*, S1pl; *e-*, 2sg):

- (2)
- | | | | | | |
|-----------------|---------------|-----------------|----------------|---------------|-------------|
| <i>*ku-e-</i> | 'I-you (sg.)' | <i>*ci-e-</i> | 'we-you (sg.)' | \Rightarrow | <i>eci-</i> |
| <i>*ku-eci-</i> | 'I-you (pl.)' | <i>*ci-eci-</i> | 'we-you (pl.)' | | |

I assume that this is the effect of two different constraints, one suppressing subject agreement in $1 \rightarrow 2$ forms, and a second one that disallows number expression by *e*: [+2-pl] and effects that 2sg object agreement is also expressed by *eci*: [+2]. Note that I take *eci* to be an underspecified 2nd person marker, not a 2nd person plural affix since it expresses 2nd person for plural *and* singular arguments albeit in partly different contexts.[‡] The formal nature of the

[‡] An anonymous reviewer suggests that the appearance of *eci-* in $1 \rightarrow 2$ forms where both arguments are singular could be analyzed as plural agreement since agreement is with more than one arguments. This would be parallel to plural agreement with two coordinated singular NPs as in *John and Mary laugh*. I think that this analysis is problematic for two reasons: First, contrary to the coordination case, this would imply agreement with a unit which is not a syntactic constituent (subject + object). Second, this analysis fails to explain why plural marking would be restricted to $1 \rightarrow 2$ configurations. In cases where coordinated NPs trigger plural agreement, this extends to my knowledge always to all person combinations.

constraint which suppresses number distinction for 2nd person agreement of the forms in (2) is the topic of this paper.

3. Distributed optimality

In Distributed Optimality (Trommer 2001,2002b), syntactic operations manipulate abstract heads without phonological features. As in Distributed Morphology (Halle and Marantz 1993), Morphology constitutes an independent module of the grammar that takes wordlike units from the output of syntax as its input and assigns to them strings of vocabulary items (VIs), pairings of underspecified syntactic feature structures and phonological matrices. In contrast to Distributed Morphology, morphological spellout of syntax is not based on language-specific rules but happens by evaluating a language-specific ranking of a universal set of morphological constraints.

For *eci-un-kore* in (1), I assume the input [+Nom+2+pl][+Acc+1+pl] (omitting the verb). PARSE [F] is violated by each input feature not realized in the output, and $L \Leftarrow [+2]$ is an alignment constraint which requires 2nd person affixes to be maximally leftwards:

(3) **Input:** [+Nom+2+pl]₁ [+Acc+1+pl]₂

	$L \Leftarrow [+2]$	PRS [F]
☞ a. <i>eci</i> :[+2] ₁ - <i>un</i> :[+Acc+1+pl] ₂ -		**
b. <i>un</i> :[+Acc+1+pl] ₂ - <i>eci</i> :[+2] ₁ -	*!	**
c. <i>un</i> :[+Acc+1+pl] ₂ -		***!
d. <i>eci</i> :[+2] ₁ -		***!***

Note that some violations of PARSE [F] cannot be avoided since there are no VIs expressing e.g. [+Acc+2+pl]. Following Woolford (2003) and Gerlach (1998), I analyze suppression of the [+1] affix for 1 → 2 forms as the effect of two alignment constraints ranked above faithfulness (here PARSE [F]):

(4) **Input:** [+Nom+1-pl]₁ [+Acc+2+pl]₂

	PRS PER ^{[+2]/[+1]}	$L \Leftarrow$ [+Nom]	$L \Leftarrow$ [+2]	PRS [F]
a. <i>eci</i> :[+2] ₂ - <i>ku</i> :[+1+Nom] ₁ -		*!		**
b. <i>ku</i> :[+1+Nom] ₁ - <i>eci</i> :[+2] ₂ -			*!	**
c. <i>ku</i> :[+1+Nom] ₁ -	*!			**
☞ d. <i>eci</i> :[+2] ₂ -				**

PARSE PER^{[+2]/[+1]} belongs to the family of relativized PARSE constraints (Trommer 2002b,2003), and is to be read as: "If there are adjacent [+2] and [+1] heads in the input, then realize the person feature ([PER]) of the [+2] head. Relativized PARSE constraints are related to universal prominence hierarchies by the schema in (5):

(5) If $A_1 \dots A_n$ are distinct from $B_1 \dots B_n$, and $A_i \geq B_i$ on a scale S_i
 $(1 \leq i \leq n)$, then there is a constraint PARSE [AGR]^{[A₁ ... A_n] / [B₁ ... B_n]}

Given the scales in (6) which are justified by extensive crosslinguistic evidence, we get particular constraints as in (7). "[+high]" stands for the highest argument that agrees with the verb, i.e., transitive subject or intransitive object, "[-high]" corresponds to intransitive subject or transitive object.

- (6)
- a. $\left\{ \begin{array}{l} [+1] \\ [+2] \end{array} \right\} > [+3]$ b. $[+high] > [+low]$
- c. $[-marked] > [+marked]$ (Nominative/Absolutive $>$ Ergative/Accusative)

(7a,b) encode that agreement with local person is preferred over agreement with 3rd person, (7c) captures the preference for subject agreement. Since $[+1]$ and $[+2]$ are not ranked, there are antagonistic constraints for verbs with $[+1]$ and $[+2]$ agreement (7d,e). Actual preference depends on the language-specific ranking. (7e) is the constraint from (4) and by assumption ranked higher than (7d) in Ainu. §

- (7)
- a. PARSE [PER]^{[+1]/[+3]} b. PARSE [PER]^{[+2]/[+3]}
- c. PARSE [PER]^{[+high]/[+low]}
- d. PARSE [PER]^{[+1]/[+2]} e. PARSE [PER]^{[+2]/[+1]}

Note that we still have no account for the fact that number is neutralized in $1 \rightarrow 2$ forms since PARSE [F] should prefer e:[+2-pl] over eci:[+2] for inputs of the form $[+Nom+1+/-pl]$ $[+Acc+2-pl]$, and no other constraint disfavors e:[+2-pl]. I will treat this problem under a crosslinguistic perspective on participant reduction.

4. Participant reduction crosslinguistically

As Noyer (1992) observes, participant reduction is widespread involving considerable crosslinguistic variation, especially inside the Tanoan Tiwa family, as to which number contrasts are neutralized when both arguments are SAPs. Thus in Nunggubuyu number of 1st person arguments is deleted. in Arizona Tiwa, all number contrasts are suppressed, in Rio Grande Tiwa only number of a 1st person subject is preserved, and in Northern Tiwa only number of a 2nd person object. In Southern Tiwa only number features of objects are preserved. Swahili is a language where all number contrasts are preserved. These constellations are summarized in (8) where "†" stands for neutralization and "👉" for retention of the number contrast in the boldfaced category of the respective row:

(8)

	Nunggubuyu	N. Tiwa	S. Tiwa	A. Tiwa	R.G. Tiwa	Swahili
1:2	†	†	†	†	👉	👉
1:2	👉	👉	👉	†	†	👉
2:1	👉	†	†	†	†	👉
2:1	†	†	👉	†	†	👉

The particular challenge participant reduction poses for a theory of morphosyntax is how to account for the basic tendency to suppress number features while capturing the degree to which this happens in single languages. Noyer (1992) who approaches this phenomenon by inviolable constraints has to assume a family of slightly different participant reduction constraints for different languages which fails to capture the common principle in all of them. I formulate the crosslinguistic tendency to syncretize number contrasts in agreement when both arguments are non-third person in (9):||

§ A similar preference for $[+2]$ over $[+1]$ prefixes is found in the Algonquian language Menominee (cf. Trommer 2002a).

|| A possible functional explanation for Participant Uniqueness might invoke the fact that speaker and hearer in a discourse are normally uniquely identified which makes number marking superfluous. However this does not

- (9) **Participant Uniqueness (P.U.):** For two adjacent [-3] agreement heads in the input, number should not be expressed in the output

(9) ranked above PARSE [F] accounts for Arizona Tiwa, the opposite ranking for Swahili. (10) and (11) show this for the input [+Nom+1+pl]₁ [+Acc+2+pl]₂.

(10) **Swahili**

	PRS [F]	P.U.
a. [+1] ₁ -[+2] ₂	*!*	
b. [+1+pl] ₁ -[+2] ₂	*!	*
c. [+1] ₁ -[+2+pl] ₂	*!	*
☞ d. [+1+pl] ₁ -[+2+pl] ₂		**

(11) **A. Tiwa**

	P.U.	PRS [F]
☞ a. [+1] ₁ -[+2] ₂		**
b. [+1+pl] ₁ -[+2] ₂	*!	*
c. [+1] ₁ -[+2+pl] ₂	*!	*
d. [+1+pl] ₁ -[+2+pl] ₂	*!*	

The languages "in-between", i.e., with partial neutralization of number, can be captured by relativized PARSE constraints, this time referring to number, instead of person and ranked above P.U.:

- (12) Nunggubuyu PRS [NUM]^{[+2]/[+1]} >> P.U. >> PRS ...
S. Tiwa PRS [NUM]^{[-marked]/[+marked]} >> P.U. >> PRS ...
R.G. Tiwa PRS [NUM]^{[+1+high]/[+2+low]} >> P.U. >> PRS ...
N. Tiwa PRS [NUM]^{[+2-marked]/[+1+marked]} >> P.U. >> PRS ...

Note that [+marked] and [-marked] refer to ergative and absolutive case in Northern and Southern Tiwa. (13) and (14) shows how the correct distribution of number marking in Northern Tiwa is derived. PRS [NUM]^{[+2-marked]/[+1+marked]} favors retention of the plural feature for the 2nd person argument in (13) since this is absolutive ([-marked]) but not for the 2nd person ([-marked]) ergative in (14). All other relativized PARSE constraints for number are assumed to be ranked below PRS [F] and are hence irrelevant.

(13) **N. Tiwa, Input:** [+Erg +1 +pl]₁ [+Abs +2 +pl]₂

	PRS [NUM] ^{[+2-marked]/[+1+marked]}	P.U.	PRS [F]
a. [+1] ₁ -[+2] ₂	*!		**
b. [+1+pl] ₁ -[+2] ₂	*!	*	*
☞ c. [+1] ₁ -[+2+pl] ₂		*	*
d. [+1+pl] ₁ -[+2+pl] ₂		**!	

explain why number marking for agreement with 1st and 2nd person arguments is retained in these languages with intransitive verbs.

(14) **N. Tiwa, Input:** [+Erg +2 +pl]₁ [+Abs +1 +pl]₂

	PRS [NUM] ^{[+2-marked]/[+1+marked]}	P.U.	PRS [F]
☞ a. [+1] ₂ -[+2] ₁			**
b. [+1+pl] ₂ -[+2] ₁		*!	*
c. [+1] ₂ -[+2+pl] ₁		*!	*
d. [+1+pl] ₂ -[+2+pl] ₁		*!*	

In the next section, I show how the crosslinguistic account to participant reduction carries over to Colloquial Ainu and allows a complete analysis of the data from section 2.

5. Ainu participant reduction revisited

Since in Ainu the subject in 1 → 2 forms is completely suppressed, it is unclear whether P.U. applies to [+1] subjects, but we know from (1) that it does not apply to [+1] objects and [+2] subjects. Thus I assume that relativized PARSE constraints generally retain number for [+1] arguments and [+2] subjects but not for [+2] objects, while PARSE [NUM]^{[+1]/[+2]} is overridden by L ⇔ [+2] and L ⇔ [+Nom] which cause the dropping of the 1st person prefix. Thus we get the ranking in (15):

$$(15) \quad \left\{ \begin{array}{l} L \Leftrightarrow [+2] \\ L \Leftrightarrow [+Nom] \end{array} \right\} \gg \left\{ \begin{array}{l} \text{PARSE [NUM]}^{[+1]/[+2]} \\ \text{PARSE [NUM]}^{[+2+high]/[+1+low]} \end{array} \right\} \gg \text{P.U.}$$

(16) shows the derivation of the form *eci-kore* for ‘we give you (sg.)’. Note that PRS NUM^{[+2+high]/[+1+low]} is never violated because there is neither a [+2+high] nor a [+1+low] head in the input, and that P.U. is only relevant here since it prefers *eci*:[+2]₂ over *e*:[+2-pl]₂.

(16) **Input:** [+Nom+1+pl]₁ [+Acc+2-pl]₂

	PRS PER ^{[+2]/[+1]}	L ⇔ [+Nom]	L ⇔ [+2]	PRS NUM ^{[+1]/[+2]}	PRS NUM ^{[+2+high]/[+1+low]}	P.U.	PRS [F]
a. <i>e</i> :[+2-pl] ₂ - <i>ci</i> :[+1+Nom+pl] ₁ -		*!				**	**
b. <i>eci</i> :[+2] ₂ - <i>ci</i> :[+1+Nom+pl] ₁ -		*!				*	**
c. <i>ci</i> :[+1+Nom+pl] ₁ - <i>eci</i> :[+2] ₂ -			*!			*	**
d. <i>ci</i> :[+1+Nom+pl] ₁ - <i>e</i> :[+2-pl] ₂ -			*!			*	**
e. <i>ci</i> :[+1+Nom+pl] ₁ -	*!					*	**
☞ f. <i>eci</i> :[+2] ₂ -				*			**
g. <i>e</i> :[+2-pl] ₂ -				*		*!	**

(17) shows how *e-en-kore*, ‘you (sg.) give me’ is derived. Here PRS NUM^{[+2+high]/[+1+low]} gets crucial and ensures preference of *e*:[+2-pl]₁-*en*:[+1+Acc-pl]₂- over *eci*:[+2]₁-*en*:[+1+Acc-pl]₂-:

(17) **Input:** [+Nom+2-pl]₁ [+Acc+1-pl]₂

	PRS PER ^{[+2]/[+1]}	L ⇄ [+Nom]	L ⇄ [+2]	PRS NUM ^{[+1]/[+2]}	PRS NUM ^{[+2+high]/[+1+low]}	P.U.	PRS [F]
a. eci:[+2] ₁ -en:[+1+Acc-pl] ₂ -					*!	*	**
☞ b. e:[+2-pl] ₁ -en:[+1+Acc-pl] ₂ -						**	*
c. en:[+1+Acc-pl] ₁ -eci:[+2] ₂ -			*!		*	*	**
d. en:[+1+Acc-pl] ₁ -e:[+2-pl] ₂ -			*!			**	*
e. en:[+1+Acc-pl] ₁ -	*!				*	*	***
f. eci:[+2] ₂ -				*!	*		*****
g. e:[+2-pl] ₂ -				*!		*	*****

6. The formal nature of participant reduction

While P.U. as formulated in (9) captures the crosslinguistic tendency that number features are suppressed in transitive verbs having only SAP arguments, it is not a possible constraint in standard OT, since it refers to input features while not being a faithfulness constraint. In Distributed Optimality (Trommer 2001), it falls under the category of "Impoverishment constraints", i.e., two-level markedness constraints marking the realization of certain features given a specific input. For some of the languages in (8), (9) could be reformulated as (18) which refers only to output structures:

(18) A [-3] VI should not be specified [+pl] in a form with another [-3] VI

But (18) does not work for Ainu 1 → 2 forms, since it cannot favor (19a) over (19b), where there is no overt [+1] affix:

- (19) a. eci:[+2]-kore
b. *e:[+2-pl]-kore

Transderivational constraints (e.g. Benua 1997) might seem to be an alternative to constraints which refer to input features. Thus P.U. could be formulated like this:

(20) Transitive forms with two [-3] heads should have equal number specifications.

But as (9), (20) has to refer to the morphological input, since the forms in (19) cannot otherwise be identified as relevant forms. Indeed *e:[+2-pl]-kore* is grammatical with the interpretations "you (sg.) give" or "he gives you". The only way to maintain *e:[+2-pl]-kore* as the correct form for these meanings while disfavoring the very same form for 1 → 2 forms is to refer to the input features. Note a crucial difference to transderivational constraints in phonology: In phonology the information that two output forms are morphologically related does not strictly follow from the phonological input. However, in morphological spellout the input features are just the same features that define paradigmatic relatedness. To formalize a

constraint like (20) one has to refer to the morphological input of at least two forms. Thus, the Distributed Optimality version of P.U. is actually more restrictive than a transderivational account, since it refers only to input features, but not to output forms and input features of related forms, while the transderivational version refers to both.

References

- Benua L 1997 *Transderivational Identity: Phonological Relations between words*. (PhD thesis, University of Massachusetts)
- Gerlach B 1998 Optimale Klitiksequenzen *ZfS*, **17** 35–91
- Halle M and Marantz A 1993 Distributed Morphology and the pieces of inflection In Hale K and Keyser S J, editors, *The View from Building 20*, pages 111–176 (Cambridge MA: MIT Press)
- McCarthy J and Prince A 1994 The emergence of the unmarked: Optimality in prosodic morphology In *NELS* 24, pages 333–379 Amherst
- Noyer, R 1992 *Features, Positions and Affixes in Autonomous Morphological Structure* (PhD thesis, MIT)
- Shibatani M 1990 *The Languages of Japan* (Cambridge: Cambridge University Press)
- Trommer J 2001 *Distributed Optimality* (PhD thesis, University of Potsdam)
- 2002a *Hierarchy-based competition* Ms., available under <http://www.ling.uni-osnabrueck.de/trommer/hbc.pdf>
- 2002b Modularity in OT-morphosyntax In: Fanselow G and Féry C, editors, *Resolving Conflicts in Grammar: Optimality Theory in Syntax, Morphology and Phonology* (Special Issue 11 of Linguistische Berichte)
- 2003 Direction marking as agreement (to appear in: Junghanns U and Szucsich L, editors, *Syntactic Structures and Morphological Information*)
- Woolford E 2003 Clitics and agreement in competition: Ergative cross-referencing patterns In: Carpenter A, Coetzee A, and de Lacy P, editors, *Papers in Optimality Theory II*, (volume 26 of University of Massachusetts Occasional Papers)

Variations in OT learning

Learning OT Constraint Rankings Using a Maximum Entropy Model

Sharon Goldwater and Mark Johnson

Department of Cognitive and Linguistic Sciences, Brown University
{sharon_goldwater, mark_johnson}@brown.edu

Abstract. A weakness of standard Optimality Theory is its inability to account for grammars with free variation. We describe here the Maximum Entropy model, a general statistical model, and show how it can be applied in a constraint-based linguistic framework to model and learn grammars with free variation, as well as categorical grammars. We report the results of using the MaxEnt model for learning two different grammars: one with variation, and one without. Our results are as good as those of a previous probabilistic version of OT, the Gradual Learning Algorithm (Boersma, 1997), and we argue that our model is more general and mathematically well-motivated.

1. Introduction

One of the requirements of any successful linguistic theory is to provide an explanation of how the learner acquires the language-specific knowledge required by the theory. Optimality Theory (Prince and Smolensky, 1993) is dominant in phonology in part because there are algorithms for learning constraint rankings (Tesar and Smolensky, 1993; Pulleyblank and Turkel, 1996; Prince and Tesar, 1999). Unfortunately, most existing OT learning algorithms have two major problems. First, they are not designed to learn from noisy training data, and generally will not converge when presented with it. Second, because they learn a single OT constraint ranking, they cannot model grammars containing free variation, where a single input form has more than one grammatical output form. (This is a limitation of OT itself, rather than a weakness of the learning procedures.) In this paper, we concern ourselves with addressing these problems. In particular, we propose that a complete model of phonology and its associated learning algorithm should be able to

- learn from a corpus of real, potentially noisy, data,
- account for free variation as well as categorical distinctions,
- account for effects caused by cumulative constraint violations, and
- generalize to examples not seen in the training data.

There have been various attempts to adapt the OT model in some way to explain free variation, including floating constraints (Nagy and Reynolds, 1997), free ranking of constraints within strata (Anttila, 1997b), and strictness bands (Hayes, 2000). One of the more successful models to date is the probabilistic model proposed by Boersma (1997) and its associated learning algorithm, the Gradual Learning Algorithm. By moving away from the discrete domain of standard OT, the Gradual Learning Algorithm is able to learn from noisy input, and can accurately reproduce grammars with free variation. However, as

This research was supported in part with funding from the National Institute of Mental Health Grant #1R01MH60922-01A2.

Keller and Asudeh (2002) have pointed out, the GLA is unable to account for cumulativity effects. Keller’s own model, Linear Optimality Theory (Keller, 2000), is designed to account for cumulativity effects, but learns only from acceptability judgment data, not from actual linguistic forms.

In this paper we present a different OT-inspired model of constraint-based phonology, the Maximum Entropy model. This model is in fact a very general statistical model that has been used in many domains and whose mathematical properties are well known. Like the GLA, this model is probabilistic, making it resistant to noise, and seeks to reproduce the distribution of output forms in a training corpus, thus modeling free variation. Like Linear Optimality Theory, the MaxEnt model treats constraints as additive, thus accounting for cumulativity effects.

The connection between OT and Maximum Entropy models used in this paper has been discussed before in Eisner (2000) and Johnson (2002). The estimation procedure or learning method used in this paper is described in detail in Johnson et al. (1999), which also contains statistical consistency results. Johnson (2002) uses the same estimation procedure to learn constraint rankings for OT Lexical Functional Grammars.

The remainder of this paper is organized as follows: We first present the MaxEnt model and its application to constraint-based phonology. We report experimental results similar to those of the GLA on both categorical (no free variation) and stochastic (free variation) training data. We then discuss the question of generalization, explain why it cannot be tested using the kinds of problems presented here, and discuss how we can test for it in future work. Finally, we argue that the MaxEnt model is more general and mathematically simpler than the GLA.

2. The Maximum Entropy Model

Maximum Entropy or log-linear models are a very general class of statistical models that have been applied to problems in a wide range of fields, including computational linguistics. Logistic regression models, exponential models, Boltzmann networks, Harmonic grammars, probabilistic context free grammars, and Hidden Markov Models are all types of Maximum Entropy models. Maximum Entropy models are motivated by information theory: they are designed to include as much information as is known from the data while making no additional assumptions (i.e. they are models that have as high an entropy as possible under the constraint that they match the training data). Suppose we have some conditioning context x and a set of possible outcomes $\mathcal{Y}(x)$ that depend on the context. Then a Maximum Entropy model defines the conditional probability of any particular outcome $y \in \mathcal{Y}(x)$ given the context x as:

$$\Pr(y|x) = \frac{1}{Z(x)} \exp\left(\sum_{i=1}^m w_i f_i(y, x)\right), \text{ where} \quad (1)$$

$$Z(x) = \sum_{y \in \mathcal{Y}(x)} \exp\left(\sum_{i=1}^m w_i f_i(y, x)\right)$$

In these equations, $f_1(y, x) \dots f_m(y, x)$ are the values of m different features of the pair (y, x) , the w_i are parameters (weights) associated with those features, and $Z(x)$ is a normalizing constant obtained by summing over all possible values that y could take on in the sample space $\mathcal{Y}(x)$. In other words, the log probability of y given x is proportional to a linear combination of feature values, $\sum_{i=1}^m w_i f_i(y, x)$.

In the MaxEnt models considered here, x is an input phonological form, $\mathcal{Y}(x)$ is the set of candidate output forms (i.e., \mathcal{Y} is the Gen function) and $y \in \mathcal{Y}(x)$ is some particular candidate output form. For an Optimality Theoretic analysis with m constraints $C_1 \dots C_m$, we use a Maximum Entropy model with m features, and let the features correspond to the constraints.

Thus the feature value $f_i(y, x)$ is the number of violations of constraint C_i incurred by the input/output pair (y, x) . We can think of the parameter weights w_i as the ranking values of the constraints.

Note that this Maximum Entropy model of phonology differs from standard Optimality Theory in that constraint weights are additive in log probability. As a result, many violations of lower-ranked constraints may outweigh fewer violations of higher-ranked constraints. This is a property shared by the recent Linear Optimality Theory (Keller, 2000), as well as the earlier theory of Harmonic Grammar (Legendre et al., 1990), on which OT is based.¹ The property of additivity makes the MaxEnt model more powerful and less restrictive than standard OT. When there is sufficient distance between the constraint weights and a finite bound on the number of constraint violations, the MaxEnt model simulates standard OT (see Johnson (2002) for an explicit formula for the weights). The model can therefore account for categorical grammars where a single violation of a highly ranked constraint outweighs any number of violations of lower ranked constraints. However, by assigning closely spaced constraint weights, the MaxEnt model can also produce grammars with variable outputs, or gradient grammaticality effects caused by cumulative constraint violations (Keller, 2000; Keller and Asudeh, 2002). The GLA is able to model grammars with free variation, but, like standard OT, cannot account for these cases of cumulative constraint violations.

Given the generic Maximum Entropy model, we still need to find the correct constraint weights for a given set of training data. We can do this using maximum likelihood estimation on the conditional likelihood (or *pseudo-likelihood*) of the data given the observed outputs:

$$\text{PL}_{\bar{w}}(\bar{y}|\bar{x}) = \prod_{j=1}^n \text{Pr}_{\bar{w}}(Y = y_j | x(Y) = x_j) \quad (2)$$

Here, $\bar{y} = y_1 \dots y_n$ are the winning output forms for each of the n training examples in the corpus, and the x_j are the corresponding input forms. So the pseudo-likelihood of the training corpus is simply the product of the conditional probabilities of each output form given its input form. As with ordinary maximum likelihood estimation, we can maximize the pseudo-likelihood function by taking its log and finding the maximum using any standard optimization algorithm. In the experiments below, we used the Conjugate Gradient algorithm (Press et al., 1992).

To prevent overfitting the training data, we introduce a regularizing bias term, or prior, as described in Johnson et al. (1999). The prior for each weight w_i is a Gaussian distribution with mean μ_i and standard deviation σ_i that is multiplied by the psuedo-likelihood in (2). In terms of the log likelihood, the prior term is a quadratic, so our learning algorithm finds the w_i that maximize the following objective function:

$$\log \text{PL}_{\bar{w}}(\bar{y}|\bar{x}) - \sum_{i=1}^m \frac{(w_i - \mu_i)^2}{2\sigma_i^2} \quad (3)$$

For simplicity, the experiments reported here were conducted using the same prior for each constraint weight, with $\mu_i = 0$ and $\sigma_i = \sigma$. (For possible theoretical implications of this choice, see Section 4.1.) Informally, this prior specifies that zero is the default weight of any constraint (which means the constraint has no effect on the output), so we can vary how closely the model fits the data by varying the standard deviation, σ . Lower values of σ give a more peaked prior distribution and require more data to force the constraint weights away from zero, while higher values give a better fit with less data, but may result in overfitting the data. In particular, multiplying the number of training examples by a factor of r (while

¹ In fact, the Harmony function from Harmonic Grammar is simply $\log \text{Pr}(y|x)$ in (2) (Smolensky and Legendre, 2002).

Constraint	Weight
*RTRHI	33.89
PARSE[RTR]	17.00
GESTURE[CONTOUR]	10.00
PARSE[ATR]	3.53
*ATRLO	0.41

Table 1. Constraint weights learned by MaxEnt model

keeping the empirical distribution fixed) will yield the same result as reducing σ by a factor of \sqrt{r} . In other words, if we vary n and σ but hold $n\sigma^2$ constant, the parameter weights learned by the MaxEnt model will be the same.

3. Experimental Results

We ran experiments on two different sets of data, one categorical and one stochastic. Both datasets are available as part of the Praat program (Boersma and Weenink, 2000). In this section, we describe our experimental results and compare them to the results of the GLA on the same datasets, as reported in Boersma (1999) and Boersma and Hayes (2001).

3.1. Learning a Categorical Grammar

For this experiment, we used the Wolof tongue-root grammar described in Boersma (1999), which includes five constraints:

*RTRHI: High vowels must not have a retracted tongue root (rtr).

*ATRLO: Low vowels must not have an advanced tongue root (atr).

PARSE[RTR]: If an input segment is [rtr], it must be realized as [rtr] in the output.

PARSE[ATR]: If an input segment is [atr], it must be realized as [atr] in the output.

GESTURE[CONTOUR]: Do not change from [rtr] to [atr], or vice versa, within a word.

There are 36 input forms provided with this grammar, each of which is paired with a winning output form and three losing candidates. Boersma (1999) reports the results of a sample run of the GLA on this set of data. The algorithm was presented with 10,000 training examples (uniformly distributed among the 36 input forms) with a plasticity of 1.0 and evaluation noise of 2.0,² and learned the following ranking:

*RTRHI»PARSE[RTR]»GESTURE[CONTOUR]»PARSE[ATR]»*ATRLO

The learned ranking values are sufficiently far apart that the noisy evaluation hardly ever reranks the constraints, giving an error rate below 0.2 percent for all input forms.

We tested the MaxEnt model using various values of $n\sigma^2$, with training data uniformly distributed among the 36 input forms. Like Boersma (1999), we tested the accuracy of the learner on these same 36 input forms. (We discuss ways to test the generalization abilities of the two algorithms in Section 4.3.) In Table 1, we show the constraint weights learned by the MaxEnt model with $n\sigma^2$ at approximately 1,200,000. With these weights, the average error rate over all input forms is 0.07 percent, and the maximum error rate for any input form is 0.19 percent (comparable to the GLA). If we increase $n\sigma^2$, the error rates drop essentially to

² See Boersma and Hayes (2001) for a description of the GLA, including an explanation of the plasticity value and evaluation noise.

zero. Note that the constraint weights learned by the MaxEnt model have the same relative ranking as those learned by the GLA and are spaced out at roughly exponential intervals. This sort of exponential pattern of constraint weights is exactly the pattern that, in the limit, gives rise to the strict domination of Optimality Theory (Johnson, 2002).

3.2. Learning a Stochastic Grammar

For this experiment, we used the data on Finnish genitive plurals described in Boersma and Hayes (2001) (henceforth B&H). This set of data was originally collected by Anttila (1997a; 1997b) from a large text corpus.

In Finnish, there are two possible genitive plural endings—a weak ending (usually /-jen/) and a strong ending (usually /-iden/). Some stems allow only one of the two endings (e.g. *kameroiden*/**kamerojen* ‘camera’, *kalojen*/**kaloiden* ‘fish’), while others are acceptable with either ending (e.g. *naapurien*/*naapureiden* ‘neighbor’). Among the stems that allow both endings, there are differences in the degree to which one ending is preferred over the other, as measured by corpus frequency. Anttila argues that the use of the weak or strong ending is determined entirely by the phonological properties of the stem. He proposes a number of possible constraints in his analysis, of which B&H use 11. Since our aim is to compare the performance of our algorithm to the results in B&H, we use these same 11 constraints:

C_1 (STRESS-TO-WEIGHT): Stressed syllables must be heavy.

C_2 (WEIGHT-TO-STRESS): Heavy syllables must bear stress.

C_3, C_4, C_5 (* \acute{I} , * \acute{O} , * \acute{A}): No stressed syllables with underlying high/mid/low vowels.³

C_6, C_7, C_8 (* \check{I} , * \check{O} , * \check{A}): No unstressed syllables with underlying high/mid/low vowels.

C_9 (*H.H): No consecutive heavy syllables.

C_{10} (*L.L): No consecutive light syllables.

C_{11} (*LAPSE): No consecutive unstressed syllables.

The data set in B&H contains 5698 tokens, which comprise all genitive plurals of stems ending in light syllables. (Stems ending in heavy syllables require the strong ending and exhibit no variation, so B&H excludes them from the test of stochastic learning.) The tokens are divided into 22 classes depending on the phonological structure of the stem. For each of these classes, the pattern of constraint violations for the winning candidate and the losing candidate is different. Table 2 shows examples of four words from different stem classes and their patterns of constraint violations.

B&H’s characterization of the data is misleading, however. Although each of the 22 classes has a different pattern of constraint violations, the GLA does not consider these patterns directly during the learning process. Rather, it learns from the pattern of *differences* between the violations of the winning output and its corresponding losing candidate. Table 3 shows the pattern of differences for each of the stems in Table 2, obtained by subtracting the vector of constraint violations for the winning candidate from that of the losing candidate. Here, we see that from the algorithm’s point of view, stems like ‘naapuri’ and ‘ministeri’ do not belong to different classes at all. Reanalyzing B&H’s classes in this way, it turns out that in fact there are only eight different classes of stems for which distributions must be learned. Since our algorithm, like the GLA, considers only differences in violations between winning and losing candidates, we consider only these eight collapsed classes in reporting our results.

Table 4 compares the results of the GLA and MaxEnt models on this data set. The “Tokens” column shows the number of tokens in each class, and the “% Majority” column

³ By “underlying vowels”, Anttila means vowels in the stem.

Word	Candidates	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}
kala	ká.lo.jen	1	1	0	0	0	0	0	1	0	1	1
	ká.loi.den	1	2	0	0	0	0	0	1	1	0	1
naapuri	náa.pu.ri.en	0	1	0	0	0	1	0	0	0	1	2
	náa.pu.rèi.den	0	1	1	0	0	0	0	0	1	0	0
ministeri	mí.nis.te.ri.en	1	2	0	0	0	1	0	0	0	1	3
	mí.nis.te.rèi.den	1	2	1	0	0	0	0	0	1	0	1
maailma	máa.il.mo.jen	0	2	0	0	0	0	0	1	1	0	2
	máa.il.mòi.den	0	2	0	0	1	0	0	0	3	0	0

Table 2. Constraint violation patterns of four of B&H’s classes, with example words

Word	Differences in Constraint Violations											
kala	0	1	0	0	0	0	0	0	1	-1	0	0
naapuri	0	0	1	0	0	-1	0	0	1	-1	-2	0
ministeri	0	0	1	0	0	-1	0	0	1	-1	-2	0
maailma	0	0	0	0	1	0	0	-1	2	0	-2	0

Table 3. Some of B&H’s classes are not distinct

Class	Tokens	% Majority	GLA	MaxEnt
1	1097	100	99.5	99.6
2	1000	100	100.0	100.0
3	923	100	100.0	100.0
4	873	70.7	69.5	69.4
5	821	98.4	100	99.8
6	457	99.6	99.4	98.0
7	436	82.1	81.6	80.5
8	91	50.5	58.0	55.3

Table 4. Results of the GLA and MaxEnt on the Stochastic Grammar

shows the percentage of output forms of that class in the training data belonging to the majority output. For example, in class 2, 100% of the output forms belong to the majority output (in this case, /-iden/), whereas in class 6, the outputs are split 70/30 (the more common ending in this case happens to be /-jen/). The “GLA” and “MaxEnt” columns show the percentage of forms produced by these algorithms that match the majority output forms in the training data. The MaxEnt results are for $n\sigma^2 = 569,800$. The GLA results are those reported in B&H, and reflect an average taken over 100 separate runs of the algorithm. During each run, the algorithm was presented with 388,000 training examples. The distribution of input forms in training was according to their empirical frequencies in the corpus, as was the distribution of output forms for each input. The training examples were presented in five groups. The initial plasticity was set to 2.0, but was reduced after each group of examples, to a final value of 0.002. The noise value began at 10.0 for the first group of training examples, and was set to 2.0 for the remaining examples. In their paper, Boersma and Hayes argue that reducing the plasticity corresponds to the child’s decreasing ability to learn with age, but give no justification for the change in noise level. In any case, it is not clear how they chose the particular training schedule they report, or whether other training schedules would yield

significantly different results. We discuss these points further in Section 4.4.

4. Discussion

In this section, we discuss some of the theoretical implications of our work and the question of generalization. We then compare the results presented for the GLA and MaxEnt model and argue in favor of the MaxEnt model on formal and practical grounds.

4.1. The Initial State

For many applications of the MaxEnt model, the bias term in the objective function is simply a means of preventing overtraining. Here, we can interpret it on a more theoretical level as a learning bias or assumption about the initial state of acquisition. To keep our initial experiments as simple as possible, we used the same prior for each constraint weight, which corresponds to the assumption that all constraints are equally ranked in the absence of data. However, it is widely believed that in fact children’s acquisition begins with markedness constraints outranking faithfulness constraints. This situation could easily be modeled by using priors with different means for the markedness and faithfulness constraints, and setting the means for the markedness constraints to some higher value than those for the faithfulness constraints. In the absence of data, markedness would outrank faithfulness, but as data accumulated indicating otherwise, the strength of the data would overcome the prior, and the faithfulness constraints would become more important. Universal rankings could be modeled similarly by adjusting the priors on various constraints to reflect the desired universal ranking.

4.2. The Learning Path

Unlike the GLA and related approaches, our approach cleanly distinguishes the structure of the model (i.e., the MaxEnt exponential form conditional probability distribution (2) and the objective function (3) to be maximized in learning) from the details of the method(s) that can be used to actually maximize that function. This corresponds to the distinction between Marr’s *computational level*, which specifies what is to be computed, and Marr’s *algorithmic level*, which specifies the algorithms used to carry out that computation (Marr, 1982). This paper’s principal claim is that the constraint weights that maximize (3) define a conditional probability distribution (2) that is as accurate as the distributions inferred by the GLA for the cases investigated here.

Any algorithm for maximizing (3) can in principle be used to find the optimal constraint weights. We used the Conjugate Gradient algorithm because it is a well-known efficient general-purpose algorithm that works well on large systems (for other tasks we have used it with thousands of constraint weights and tens of thousands of training items), but there are a number of other algorithms that could be used instead. For example, *iterative scaling algorithms* are specialized for optimizing MaxEnt objective functions (Berger et al., 1996) but should yield the same results as obtained with the Conjugate Gradient algorithm. *Gradient ascent* is a popular but not very efficient optimization algorithm which may produce human-like learning curves, although we have not investigated this here: again, the constraint weights it converges to should be the same as the ones obtained using Conjugate Gradient.⁴ We leave for future work the question of which optimization algorithm best models the human learning path.

⁴ This discussion ignores the possibility of multiple local maxima. In fact it is possible to show that the log conditional likelihood is concave, so there is only one global maximum (Berger et al., 1996).

4.3. Generalization

In the machine learning community, it is standard practice to evaluate the generalization ability of a learning algorithm by testing on examples not seen in the training data. This is typically done by partitioning the corpus, training on, say, 90% of the data, and testing on the remaining 10%. For small data sets, this process can be repeated using the other nine possible partitions of the corpus to obtain an average test set performance. For very small data sets, the testing portion may consist of only a single data point. Keller and Asudeh (2002) suggest using exactly these methods to evaluate the generalization ability of the GLA, and at first glance, it seems that we should evaluate the MaxEnt learner in this way.

Upon reflection, however, this sort of experiment doesn't make sense for the learning problems we have seen so far. We could set aside 10% of the 5698 Finnish words for testing, but the learning algorithm doesn't see words, it only sees patterns of violations. Since all the words in the corpus fall into only eight classes of violation patterns, the learning algorithm would have already seen many examples of each class during training, and there would be no need to generalize during testing. Alternatively, we could treat the classes themselves as the data points, and perform a leave-one-out regimen. But that would be like providing a child with input that is missing all words with certain phonological characteristics, and expecting the child to be able to produce those words correctly. This is not the normal course of acquisition.

The reason there is no real generalization problem in the tasks we have seen so far is that much of the work has been done before training even begins. The small number of word classes is due to the fact that linguists have chosen a few relevant constraints by which to characterize each word. One of our stated criteria for a successful learning algorithm is the ability to generalize, but we will not be able to test this ability until we start working on more difficult problems. These would be problems with many more constraints, so that the number of possible combinations of constraint violations would be large enough that the algorithm would not see all of the possibilities during training. We are currently working on finding data for a problem of this type in order to truly test the generalization ability of the MaxEnt learner.

4.4. Comparison to the GLA

We believe there are three key features of the GLA that have caused it to become influential. First is its ability to model variation in the adult grammar. Second is the ability to model the initial state (by setting the initial rankings of faithfulness and markedness constraints to different values). Finally, in at least some cases, the GLA seems to mimic the child's learning path (Boersma and Levelt, 1999). We have shown that the MaxEnt algorithm is able to learn both categorical and stochastic grammars as accurately as the GLA. We have not yet run experiments using different priors or different learning algorithms, but we have shown that it would be easy to use these methods to model different assumptions about the initial state and the learning path.

Given the preliminary nature of our results with regard to the actual process of acquisition, why do we believe the MaxEnt model is worth pursuing as an alternative to the GLA? Our argument is twofold. First, the MaxEnt model is mathematically well-motivated, resting on principles of information theory. It has only a single parameter to set—the ratio of σ , the standard deviation of the prior, to the number of training examples (i.e. how closely the model should fit the data). The GLA, in contrast, has at least two parameters—the ratio of the plasticity value to the number of training examples, and the evaluation noise—and potentially many more, if complicated training schedules like the ones in B&H are used. There seems to be no principled way to choose the parameters for a good training schedule, nor do we know

how sensitive the results are to that choice, or whether the GLA is guaranteed to converge. In contrast, there is a clear relationship between $n\sigma^2$ and the accuracy of learning in the MaxEnt model, and many optimization algorithms that could be used, including Conjugate Gradient, have proofs of convergence.

The second advantage of the MaxEnt model is its generality. Unlike the GLA, the MaxEnt model is not designed specifically for OT, and in fact has been used in many other fields for a century since its original introduction in statistical physics. The mathematical properties of the model have been well-studied, it has been shown to be useful for learning in a variety of domains, and in general there is a wide literature available (Jelinek, 1997).

5. Conclusions

In this paper we have presented a new way of modeling constraint-based phonology using the statistical framework of the Maximum Entropy model. We have shown that this model, in conjunction with standard optimization algorithms, can learn both categorical and stochastic grammars from a training corpus of input/output pairs. Its performance on these tasks is similar to that of the GLA. We have not yet added any assumptions about the initial state or learning path taken by the MaxEnt model, but we have described how this could easily be done by changing the priors of the model or the optimization algorithm used.

In addition to these empirical facts about the MaxEnt model, we wish to emphasize its strong theoretical foundations. Unlike the GLA, which is a somewhat ad hoc model designed specifically for learning OT constraint rankings, the MaxEnt model is a very general statistical model with an information theoretic justification that has been used successfully for many different types of learning problems. The MaxEnt model also has fewer parameters than the GLA and does not require complicated training schedules. Given our positive results so far and the success of Maximum Entropy models for other types of machine learning tasks, we believe that this model is worth pursuing as a framework for probabilistic constraint-based phonology.

References

- Arto Anttila. 1997a. Deriving variation from grammar: a study of finnish genitives. In F. Hinskens, R. van Hout, and L. Wetzels, editors, *Variation, change and phonological theory*, pages 35–68. John Benjamins, Amsterdam. Rutgers Optimality Archive ROA-63.
- Arto Anttila. 1997b. *Variation in Finnish phonology and morphology*. Ph.D. thesis, Stanford Univ.
- Adam L. Berger, Vincent J. Della Pietra, and Stephen A. Della Pietra. 1996. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22(1):39–71.
- Paul Boersma and Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry*, 32(1):45–86.
- Paul Boersma and Clara Levelt. 1999. Gradual constraint-ranking learning algorithm predicts acquisition order. In *Proceedings of the 30th Child Language Research Forum*.
- Paul Boersma and David Weenink. 2000. Praat, a system for doing phonetics by computer. <http://www.praat.org>.
- Paul Boersma. 1997. How we learn variation, optionality, and probability. In *Proceedings of the Institute of Phonetic Sciences of the Univ. of Amsterdam*, volume 21, pages 43–58.
- Paul Boersma. 1999. Optimality-theoretic learning in the praat program. In *Proceedings of the Institute of Phonetic Sciences of the Univ. of Amsterdam*, volume 23, pages 17–35.

- Jason Eisner. 2000. Review of Kager: "Optimality Theory". *Computational Linguistics*, 26(2):286–290.
- Bruce Hayes. 2000. Gradient well-formedness in optimality theory. In J. Dekkers, F. van der Leeuw, and J. van de Weijer, editors, *Optimality Theory: Phonology, Syntax, and Acquisition*. Oxford University Press, Oxford.
- Frederick Jelinek. 1997. *Statistical Methods for Speech Recognition*. The MIT Press, Cambridge, Massachusetts.
- Mark Johnson, Stuart Geman, Stephen Canon, Zhiyi Chi, and Stefan Riezler. 1999. Estimators for stochastic 'unification-based' grammars. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*.
- Mark Johnson. 2002. Optimality-theoretic Lexical Functional Grammar. In Paula Merlo and Susan Stevenson, editors, *The Lexical Basis of Sentence Processing: Formal, Computational and Experimental Issues*, pages 59–74. John Benjamins, Amsterdam, The Netherlands.
- Frank Keller and Ash Asudeh. 2002. Probabilistic learning algorithms and optimality theory. *Linguistic Inquiry*, 33(2):225–244.
- Frank Keller. 2000. *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality*. Ph.D. thesis, Univ. of Edinburgh.
- Géraldine Legendre, Yoshiro Miyata, and Paul Smolensky. 1990. Harmonic grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. Technical Report 90-5, Institute of Cognitive Science, Univ. of Colorado.
- David Marr. 1982. *Vision*. W.H. Freeman and Company, New York.
- Naomi Nagy and Bill Reynolds. 1997. Optimality theory and variable word-final deletion in faetar. *Language Variation and Change*, 9:37–55.
- William Press, Saul Teukolsky, William Vetterling, and Brian Flannery. 1992. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, England, 2 edition.
- Alan Prince and Paul Smolensky. 1993. Optimality theory: Constraint interaction in generative grammar. Technical Report TR-2, Rutgers Center for Cognitive Science, Rutgers Univ.
- Alan Prince and Bruce Tesar. 1999. Learning phonotactic distributions. Technical Report TR-54, Rutgers Center for Cognitive Science, Rutgers Univ. Rutgers Optimality Archive ROA-353.
- Douglas Pulleyblank and William J. Turkel. 1996. Optimality theory and learning algorithms: The representation of recurrent featural asymmetries. In J. Durand and B. Laks, editors, *Current trends in phonology: Models and methods*, pages 653–684. Univ. of Salford.
- Paul Smolensky and Géraldine Legendre. 2002. The harmonic mind: From neural computation to optimality-theoretic grammar. Book draft.
- Bruce Tesar and Paul Smolensky. 1993. The learnability of optimality theory: An algorithm and some basic complexity results. Ms., Department of Computer Science and Institute of Cognitive Science, Univ. Of Colorado, Boulder. Rutgers Optimality Archive ROA-2.

evolOT

Software for simulating language evolution using Stochastic Optimality Theory

Gerhard Jäger

University of Potsdam
Institut für Linguistik
Postfach 601553
D-14415 Potsdam, Germany

1. Introduction

The *evolOT* program is an implementation of the iterated “Bidirectional Gradual Learning Algorithm” (BiGLA) for Stochastic Optimality Theory [2], a variant of Paul Boersma’s Gradual Learning Algorithm (GLA [1]). It takes the insights of recent work on bidirectional OT into account. Iterated BiGLA has successfully been used to derive abstract properties of natural language like iconicity, as well as empirically attested universals like the correlation between Differential Case Marking and animacy, as evolutionary stable properties. This abstract describes the algorithms that are implemented by *evolOT*. The software can be freely downloaded from <http://www.ling.uni-potsdam.de/~jaeger/evolOT>. There you will also find installation instructions and further useful information.

2. The algorithms

Paul Boersma’s GLA is an algorithm for learning a Stochastic OT grammar. It maps a set of utterance tokens—a training corpus—to a grammar that describes the language from which this corpus is drawn. As a stochastic grammar, the acquired grammar makes not just predictions about grammaticality and ungrammaticality, but it assigns probability distributions over each non-empty set of potential utterances. If learning is successful, these probabilities converge towards the relative frequencies of utterance types in the training corpus.

GLA operates on a predefined generator relation GEN that determines what qualifies as possible inputs and outputs, and which input-output pairs are admitted by the grammatical architecture. Furthermore it is assumed that a set CON of constraints is given, i.e. a set of functions which each assign a natural number (the “number of violations”) to each element of GEN.

GLA maps these components alongside with the training corpus to a ranking of CON on a continuous scale, i.e. it assigns each constraint a real number, its *rank*.

At each stage of the learning process, GLA assumes a certain constraint ranking. As an elementary learning step, GLA is confronted with an element of the training corpus, i.e. an input-output pair. The current grammar of the algorithm defines a probability distribution over possible outputs for the observed input, and the algorithm draws its own output for this input at random according to this distribution. If the result of this sampling does not coincide with the observation, the current grammar of the algorithm is slightly modified such that the observation becomes more likely and the hypothesis of the algorithm becomes less likely. This procedure is repeated for each item from the training corpus.

The algorithm contains several parameters that can be set by the user in *evolOT*, namely the number of observations, the plasticity value, the initial ranking of the constraints, and the “noise”, i.e. the standard deviation of the normal distribution N .

In *evolOT*, the training corpus is not directly supplied by the user. Instead, the user defines a frequency distribution over GEN, and the actual training corpus is generated by a random generator interpreting the relative frequencies as probabilities.

BiGLA, the bidirectional version of GLA, differs from that in two respects. First, during the generation step the algorithm generates an optimal output for the observed input on the basis of a certain constraint ranking. It is tacitly assumed that “optimal” here means “incurring the least severe pattern of constraint violations” in standard OT fashion. In BiGLA it is instead assumed that the optimal output is selected from the set of outputs from which the input is *recoverable*. The input is recoverable from the output if among all inputs that lead to this output, the input in question incurs the least severe constraint violation profile (i.e. we apply interpretive optimization). If there are several outputs from which the input is recoverable, the optimal one (in the standard sense) is selected. If recoverability is impossible, the unidirectionally optimal output is selected.

This modification can be called “bidirectional evaluation”. Besides BiGLA involves *bidirectional learning*. This means that BiGLA both generates the optimal output for the observed input, and the optimal input for the observed output. “Comparison” and “adjustment” apply both to inputs and outputs as well. Thus the pseudo-code for BiGLA is:

Initial state All constraint values are set to the *initial value*.

for ($i := 0; i < \text{NumberOfObservations}; i := i + 1$) {

Observation A training datum is drawn at random from the training corpus, i.e. a fully specified input-output pair $\langle i, o \rangle$.

Generation

- For each constraint, a noise value is drawn from a normal distribution N and added to its current ranking. This yields the *selection point*.
- Constraints are ranked by descending order of the selection points. This yields a linear order of the constraints.
- Based on this constraint ranking, the grammar generates an optimal output o' for the input i and an optimal input i' for the output o using bidirectional evaluation.

Comparison If $i = i'$ and $o = o'$, nothing happens. Otherwise, the algorithm compares the constraint violations of the learning datum $\langle i, o \rangle$ with the self-generated pairs $\langle i, o' \rangle$ and $\langle i', o \rangle$.

Adjustment

- All constraints that favor $\langle i, o \rangle$ over $\langle i, o' \rangle$ are *promoted* by some small predefined numerical amount (“plasticity”).
- All constraints that favor $\langle i, o \rangle$ over $\langle i', o \rangle$ are *promoted* by some small predefined numerical amount (“plasticity”).
- All constraints that favor $\langle i, o' \rangle$ over $\langle i, o \rangle$ are *demoted* by the plasticity value.
- All constraints that favor $\langle i', o \rangle$ over $\langle i, o \rangle$ are *demoted* by the plasticity value.

}

evolOT allows to choose between uni- and bidirectional evaluation, and uni- vs. bidirectional learning independently. So it actually implements four different learning algorithms, GLA, BiGLA, and two mixed versions.

Depending on the OT system that is used, the training corpus and the chosen parameters, the stochastic language that is defined by the acquired grammar may deviate to a greater or lesser degree from the training language. Especially for BiGLA this deviation can be considerable. (It is perhaps misplaced to call BiGLA a “learning” algorithm; it rather describes a certain adaptation mechanism.) If a sample corpus is drawn from this language and used for another run of GLA/BiGLA, the grammar that is acquired this time may differ from the previously learned language as well.

Such a repeated cycle of grammar acquisition and language production has been dubbed the *Iterated Learning Model* of language evolution by Kirby and Hurford [3]. It is schematically depicted in figure 1.

The production half-cycle involves the usage of a random generator to produce a sample corpus from a stochastic grammar. In the *evolOT* implementation, we assume that this sample

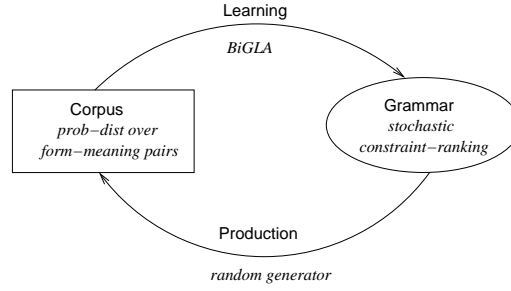


Figure 1. The Iterated Learning Model

corpus has the same absolute size than the initial corpus. Furthermore we assume that the absolute frequencies of the different *inputs* are kept constant in each cycle. What may change from cycle (“generation”) to cycle are the relative frequencies of the different outputs for each input. (I assume that the relative input frequencies are determined by extra-grammatical factors, and it is one of the main objectives of *evolOT* to model the interdependence between these factors and grammar.)

Formally put, the initial training corpus defines a frequency $\#(i)$ for each possible input i by

$$\#(i) \doteq \sum_o \#(\langle i, o \rangle)$$

where $\#(\langle i, o \rangle)$ is the number of occurrences of the utterance type $\langle i, o \rangle$ in the initial corpus. Furthermore, a given stochastic grammar G defines a probability distribution $p_G(\cdot|i)$ over the possible outputs o for each input i . Using this notation, the pseudo-code of the algorithm simulating the production step of the Iterated Learning Model can be formulated as in figure 2. I assume that there are finitely many possible inputs and outputs, which can be enumerated by as i_n, o_m etc. “*NewCorpus*” is a two-dimensional array representing the frequency distribution of the generated corpus. This means that $NewCorpus[k][l]$ is an integer representing the frequency of the pair $\langle i_k, o_l \rangle$ in the generated corpus. One cycle of learning and production represents one generation in the evolutionary process that is simulated by *evolOT*. This cycle may be repeated arbitrarily many times, i.e. over an arbitrary number of generations (which is to be fixed by the user).

```

forall k, l : NewCorpus[k][l] := 0
for (k := 0; k < NumberOfInputs; k := k + 1) {
  for (l := 0; l < #(i_k); l := l + 1) {
    o Draw an output o_n at random from the probability distribution p_G(·|i_k);
    o NewCorpus[k][n] := NewCorpus[k][n] + 1;
  }
}
  
```

Figure 2. Language production algorithm

References

- [1] Paul Boersma. *Functional Phonology*. PhD thesis, University of Amsterdam, 1998.
- [2] Gerhard Jäger. Learning constraint sub-hierarchies. The Bidirectional Gradual Learning Algorithm. In Reinhard Blutner and Henk Zeevat, editors, *Pragmatics in OT*. Palgrave MacMillan, to appear.

- [3] Simon Kirby and James R. Hurford. The emergence of linguistic structure: An overview of the Iterated Learning Model. In Domenico Parisi and Angelo Cangelosi, editors, *Simulating the Evolution of Language*, pages 121–148. Springer, Berlin, 2001.

Decision Theoretic Models of Optimality

Betsy McCall

University of Pittsburgh

Abstract. This paper examines variability in Optimality Theoretic models by considering their mathematical representations. To this end, four variations on Optimality Theory are modeled as simple Decision Theoretic utility functions that are then analyzed and compared. These versions include a strict version of OT, a version of OT that permits obligatory constraint tying, a version that permits multiple violations of individual constraints, and a stochastic model. The mathematical models help to highlight any of the theoretical difficulties in each version, as well as the power of a simple stochastic model. This paper will consider the implications that such models have for linguistic theory and for future research with respect to Universal Grammar, language acquisition, natural language processing and the dynamics of language change.

1. Introduction

Optimality Theory was first introduced to the linguistics community in 1993 in Prince and Smolensky's seminal work "Optimality Theory: Constraint Interaction in Generative Grammar". In very simple terms, Optimality Theory describes a series of ranked and interacting constraints that represent two opposing forces in language: faithfulness to some underlying representation, and well-formedness. According to the principles of Universal Grammar, all these constraints are spelled out and, while they can be reranked to accommodate acquiring a particular language, cannot be added to. This implies that there is a fixed number of N constraints.

Since the introduction of Optimality Theory, the theoretical details have been expanded by a number of people. In this paper we will not primarily be considering the different types of constraints, but the way in which constraints are violated and ranked.

Decision Theory is a science and mathematics dedicated to understanding decision-making under uncertainty. Uncertainty is present in all levels of a speaker's language understanding—in learning; in comprehension (when dealing with ambiguity resolution, for instance); and in production. By this reckoning, understanding language models through Decision Theory is a necessary approach, as Decision Theory helps us determine which strategies are reasonable when all factors affecting a situation are not known. Decision Theory allows us to convert our knowledge of the world, usually gained through statistical knowledge, into a utility function which helps us analyze future decisions based on previously acquired information.

In this paper, we will examine four different versions of Optimality Theoretic models in Decision Theoretic terms. The goal is to examine theoretical strengths and weaknesses of the different versions of Optimality Theory in order to determine which models need to be reexamined or discarded, and the nature of future research into the nature of remaining models.

2. Models

Decision Theoretic models are mathematical formulae that relate the utility of an outcome, whether it's desired or undesired and to what degree, with the expectation or probability of that outcome. For our purposes, Optimality Theory itself has taken care of this with the constraint

ranking. Rather than introducing a complex statistical utility function, we will adopt the notion of the constraint ranking, which already incorporates the notions of expectation of success or failure, and transform this into a mathematical equation that captures the violation of constraints and the relative weights of constraints. The utility functions discussed in this section are mathematically simple, yet telling.

Each of the models described are made of up two principle features: the constraints themselves and the constraint ranking. Each of the constraints, according to the theory of Universal Grammar must be listed in each speaker's grammar at birth; therefore, there cannot be an infinite number of constraints, but some finite number N of them. Each of our utility functions will be based on a summation of successfully satisfied constraints, as well as the value of success for that constraint. Because there are a finite number of constraints, we need not concern ourselves here with notions of mathematical convergence. Each of the constraints will be indicated by a variable. In the case where we do not allow multiple violations of constraints, the constraints will be given by I_{0j} . This notation indicates that each constraint is marked by an indicator variable, taking on the values of zero or one to indicate failure or success respectively, and numbered with the j -subscript as one of the N total constraints in the grammar. A constraint ranking will permute the constraints and associate them with a ranking according to their utility in a given language. The ranking itself will be given by a variable a_i that will indicate the value of a constraint associated with it being satisfied. If the constraint is satisfied the coefficient will add that much utility to the overall value of the function; if the constraint is not satisfied, the coefficient will be multiplied by a zero and no additional utility will be contributed. Each of the models considered below will rely on some variation of these basic ingredients.

2.1. *Strict Optimality Theory*

The first of the Optimality Theoretic models we will consider is a strict version of Optimality Theory at its bare bones. This version of Optimality Theory is similar to that ascribed to by John McCarthy (2002) and others. The components of this version of Optimality Theory are quite restrictive. First, although it may not be clear to an outsider how constraints are ranked when two constraints do not appear to interact, the speaker must, in fact, rank them, permitting minimal variations within a constraint ranking to produce identical grammars. Second, constraints may not accept multiple violations. Constraints are naturally only satisfied or unsatisfied—thus requiring the use of the indicator variable. Thirdly, constraint rankings, once fixed at the conclusion of language acquisition, cannot be modified and constraint rankings are impermeable, not admitting to probabilistic variation. The utility function for this strict OT is given in (1).

$$U(x) = \sum_{i=1}^n a_i I_{0j} \quad (1)$$

This equation says simply that the utility function U , operating on some element of language x , an input for instance that the grammar is analyzing for speech production, has a utility in the language equivalent to the sum of the values of the satisfied constraints. The winning candidate will be the candidate with the highest utility.

We can also take a more literal interpretation of Optimality Theory. Typically, in OT tableaux, constraint violations are marked rather than constraint successes. We can instead consider a loss function, given in (2), where constraints that are satisfied receive a loss value of zero, and constraints that are violated receive a loss value of one times the a_i value for its place in the constraint ranking. An input to the function that receives the lowest value of L is the

winning candidate. It can be shown that maximizing utility and minimizing loss are equivalent results (Berger, 1985), so that for the remainder of the paper I will primarily only be considering optimizing the relevant utility functions, although I will comment further if the correspondence between loss and utility is not obvious.

$$L(x) = \sum_{i=1}^n \alpha_i I_{0_j} \quad (2)$$

In (1) and (2), there is no specification of the values of a_i . In order to achieve the kind of constraint ranking that is described in Optimality Theory, a further specification of the values of a_i needs to be added here. So that a single constraint cannot have a lesser utility than the sum of lower ranked constraints, each coefficient in the ranking must satisfy the equation in (3).

$$\sum_{k=1}^{i-1} a_k \leq a_i \quad (3)$$

So consider, if the lowest constraint in the ranking is equivalent to a value of one, the next highest ranked constraint must be a little higher, say, $(1+\epsilon)$, where epsilon is some small amount greater than zero. The next ranked constraint must be at least this sum, and so forth. If we continue with this scheme, then if there are N constraints, the utility value of the highest ranked constraint is 2^{N-1} , and the total possible utility would be approximately 2^N . This relationship between the highest and lowest ranked constraints would be true, regardless of the scaling factor used. Since it is unlikely that for a given constraint, all constraints of lower utility will be satisfied—the higher the constraint is ranked, the less likely this becomes—we can simplify the equation in (3) so that there is just an equal sign.

Constraint interaction may also occur in strict Optimality Theory in a limited fashion through constraint conjunction. The utility model described here can be made to naturally incorporate constraint conjunction. Constraint conjunction represents a logical AND between two independent constraints. These can be derived from constraint interaction in our model by permitting multiplication of the two constraint variables that are conjoined. Both must achieve a value of one for the multiplication to be nonzero. Constraint conjunction has logical consequences for the grammar. Even if we permit only two constraints being conjoined at once, if all the possible conjunctions must be listed in Universal Grammar and not acquired during the learning process, we increase the maximal number of constraints by $N(N-1)$; i.e. the maximal utility of the grammar is now two raised to the N^2 power. If we were also to admit of language specific constraints, and expanding OT to other parts of the grammar, N becomes large very quickly and N^2 larger still. This relates directly to the problem of the infinities. Though not technically, infinite, the size of appears to be capable of growing nearly without bound.

2.2. Other constraint impermeable models

Linguists champion this kind of strict model of Optimality Theory because it is theoretically simple. Just as we can see from the mathematical representation, it requires only two relationships between the grammar and the value of an element: the ranking itself, and the relationship between the constraints and the ranking. The simpler a model is, the easier it should be to acquire and encode in UG. The drawback to the model remains in the question of whether or not it can capture all of the features of known languages and language acquisition. Thus, other models have arisen. In this section we will consider two possible variations on Optimality Theory that preserve the notions of constraint impermeability.

2.2.1. Tied constraints

A model of Optimality Theory that satisfies the second and third features of strict OT as described in §2.1, but which permits constraint tying is described here. Versions of Optimality Theory that incorporate constraint tying do so for two possible reasons. The first of these reasons was initially proposed as a possible account of producing variation within Optimality Theoretic grammars, particularly with effects such as emergence of the unmarked and context effects. The second possibility is that tied constraints can produce the effects of a logical OR within the grammar. The general utility function is given in (4). We call it U_t for ‘tied’ to distinguish it from the function for strict OT, although the equation is identical. The changes come in the way we define the coefficients that figure into the constraint ranking, given in (5).

$$U_t(x) = \sum_{i=1}^n a_i I_{0_j} \quad (4)$$

$$a_i = \begin{cases} a_k & k = i - 1 \text{ and constraint tied} \\ \sum_{k=1}^{i-1} a_k & \text{otherwise} \end{cases} \quad (5)$$

In order to achieve constraint tying, the possibility for two successive constraint weights being equal must be allowed. Equation (5) says that for most constraints, we define successive constraints as we would for strict Optimality Theory, as equal to (or greater than) the sum of all lower ranked constraints. However, this definition of the a_i ’s leaves open the possibility that a constraint may be tied in utility to the one immediately preceding it in the ranking. This formulation only tells us that a constraint, as it is added to the ranking, may be ranked equal to the previous one in the ranking. This particular model does not specify any limit on the number of constraints that may be ranked equally. To prevent this from happening, we would require another constraint, perhaps that $a_i \neq a_{i-2}$. Without this additional constraint, this clearly can be a way to reduce the maximum utility (numerical size) of the grammar by not requiring non-interacting constraints to be ranked with respect to each other, particularly for very highly constraints that are never violated in the working language, or for very low ranked constraints that are never satisfied, to be ranked equally and contribute less to the ratio between the highest ranked constraint and the lowest. Reducing the unused portions of the grammar should result in simpler computation of winning candidates by placing more emphasis on constraints that are actually decisive.

2.2.2. Multiple violations

A model of Optimality Theory that satisfies the first and third requirements described in §2.1 for strict OT, but that allows multiple violations of constraints is described in this section. Multiple violations of a constraint, or gradient effects, arise typically in certain well-formedness constraints such as those governing right- or left-headedness. If a constraint receives a violation for each syllable, for instance, as it moves into a word, it may be recorded in an analysis as receiving multiple violations if it moves beyond the first syllable. Distinguishing the accent placement, for instance between the first syllable versus the second or later syllables then, can be easily obtained from single violations, but distinguishing between second and third syllable is often achieved through allowing multiple violations. John McCarthy (2002) specifically rejects such gradient effects, but since the process is common in existing models of a wide range of phenomena, we describe it here.

Achieving multiple violations cannot be achieved through the use of an indicator variable. Rather, another variable, here labeled, z_j , is an ordinal variable. For constraints that can achieve only success or failure, nothing has changed except the label, since not all constraints need to be gradient. However, for constraints that achieve multiple violations, values of two, three, four, or whatever whole number is needed can be achieved. Our utility function now looks like (6).

$$U_{mv}(x) = \sum_{i=1}^n a_i z_j \quad (6)$$

Because we are no longer considering a simple indicator variable, we once again need to reconsider our coefficient ranking. In order to keep the strict ranking approach of previous models, we need to adjust our a_i values to accommodate multiple violations of a constraint. To guarantee that higher ranked constraints will always have a higher utility value than constraints that can have multiple violations, we need to consider the maximum utility value of the constraint in question given complete success. Our indicator variables allowed for a zero value if the constraint failed to be satisfied and a value of one if it succeeds. Now, since there are different degrees of failure, there must also be different degrees of success. Negative numbers are not allowed, so one way around this is to determine the maximum number of violations that are permitted that are still useful in the grammar. If an accent, for example, appears only on the last three syllables of a word, for instance, then three violations guarantee failure. There is no need for a fourth degree. This maximum number of violations achieves a zero value, and complete success, or no violations, receives this ordinal value in the constraint ranking. The maximum value will be something learned in language acquisition. The equation for this scheme is given in (7) and (8). We choose (7) if we wish to consider the maximum total violations (regardless of where the usefulness of such violations ends) which depends entirely upon observation, and (8) if m_j represents the maximal decisive violations associated with each constraint, something that would require a deeper understanding of the grammar. This value may indeed be one (the minimum value), and we return to strict OT if this were true for all constraints.

$$\sum_{k=1}^{i-1} a_k \max(z_j) \leq a_i \quad (7)$$

$$\sum_{k=1}^{i-1} a_k m_j \leq a_i \quad (8)$$

One of the weaknesses of such a model is that it increases the size and complexity of the grammar. The value of utilities for all successive constraints must be ranked higher to maintain the constraint ranking. The same effect might conceivably be achieved by splitting up the constraints, just as we do for place feature faithfulness and as would be done in a statistical analysis of an ordinal variable, into pieces labeled with indicator variables and ranking these successively, one after another (Kleinbaum, et al., 1998). It also forces us to establish an additional relationship between the constraints to ensure that the constraint with three violations is not ranked above the one with two violations. This approach, of course, increases our value for N. Constraint conjunction also represents a problem for constraints with multiple violations. Would conjoined constraints reduce to I_0 or maintain the gradience of the bare constraint.

It is certainly conceivable that variations on Optimality Theory exist that incorporate features of both tied constraints and multiple violations of constraints. Combining features of both constraint tying and multiple violations would not change our general utility function much, as we've seen, but would change dramatically the way in which we define our coefficients, particularly for tied, gradient constraints. I leave these variations to the imagination of the reader.

2.3. Stochastic Optimality Theory

Stochastic OT was introduced as yet another method of handling variation in a synchronic grammar. Constraint tying was proposed originally as a way of achieving variation, but in the end, this technique only permits lower ranked constraints to be the deciding factor, leading to variation which is ultimately contextual. Stochastic OT permits variation which is truly random. The mathematical model of a stochastic model of Optimality Theory is given in (9).

$$U_s(x) = \sum_{i=1}^n (a_i + b_i Y_j) I_{0_j} \quad (9)$$

The model given in (9) contains the usual features of strict OT, indicator variables for each constraint, and a coefficient a_i for the constraint ranking. The second term $b_i Y_j$ of the coefficient is the stochastic portion of the grammar, which is irrelevant if the constraint itself is not satisfied. Each Y_j represents a random variable associated with each constraint. Each Y_j takes on the value of one with probability p_j and zero with probability $(1 - p_j)$. When the random variable Y_j achieves a value of one, then the value of the coefficient b_i adds to the value of the utility function. (We assume here that the random variable is evaluated once for every input, and not once for each candidate individually.) A model for the strict version of OT can be achieved when all the p_j 's are very close to or equal to zero, as this would leave only the bare constraint ranking. However, when we change the value of some of the p_j 's, constraint permeability appears.

If the magnitude of the coefficient is free, the degree of permeability depends upon the magnitude of the coefficient of the random variable in relation to the value of the constraint itself. Values of b_i significantly smaller (or larger) than the corresponding a_i permit contextual variation with random variability, as a combination of smaller ranked constraints may combine to produce a utility greater than the single constraint alone. Values of b_i that are equal to the corresponding a_i will cause variation with the constraint ranked immediately above it. When we combine this with a p_j value equal to one, we regain the tied constraints model. The ability to recover several other models here is a strong plus for this model. This is straightforward for indicator variable constraints, but becomes more complicated for constraints that permit multiple violations, and I will not address those complications here.

In order to achieve maximum learnability, we need to gain maximum control of the theory; we would like to reduce the variation in the model to only what is needed to account for behaviour. Ideally, allowing the value of b_i to depend directly on the corresponding a_i , and $b_i + a_i \approx a_{i+1}$, so that constraint permeability is possible in only one direction, and the values of the b_i 's do not need to be acquired separately. This would permit constraint stochastic effects only with two successively ranked constraints. However, this restriction leaves open certain theoretical questions. When we consider small segments of a grammar in analyzing a particular behaviour of interest, it is not difficult to get two constraints that are varying with each other to be ranked together. The question that remains, however, is will these constraints remain

consecutively ranked when the full grammar is considered? Until complete Optimality Theoretic grammars are developed, and analyzed, that are meant to account for an entire language, complete with variation, what restrictions can be placed on the stochastic portion of this model remains to be seen.

3. Implications of the models

These mathematical models of Optimality Theory have implications for linguistic theory. Some of these implications have already been addressed above, but in this section, I would like to highlight these and others relating to some specific theoretical issues.

3.1. Universal Grammar

Universal Grammar is a central feature of modern linguistic theory. These models have a lot to say about what UG would have to contain with respect to Optimality Theory. We saw in (3), given in §2.1, how our constraint ranking must be accounted for in our utility function. Given that multiple violations and tied constraints are not a feature of this version of OT, the values of the constraint ranking for our utility function, and the utility function itself can be listed in UG. A speaker would have to acquire the permutation of constraints so that the coefficients can be associated with the correct utility values. The coefficients themselves, however, may be contained in UG since, given a fixed number of constraints, under this model the value of each coefficient would be invariant across languages. This would also be true of the stochastic model given here if we assume that the b_i 's are dependent upon the a_i 's and that $p_j=0$ is the default for all constraints initially.

On the other hand, as we've seen, if we assume that all constraints (and their binary conjunctions) are listed in UG, we have a very large grammar which to work from. This is powerful, but unwieldy. Models of UG that permit constraint learning can help to minimize the size of a grammar significantly. Humans are known to have difficulty managing small and large numbers simultaneously, so reducing a grammar to its minimal parts could be advantageous.

3.2. Language acquisition

These models address several features relevant to language acquisition. Assuming that UG conforms to the strict version of OT described in §2.1, language acquisition would be at its simplest of the four models. A speaker would have only to acquire the constraint ranking that maximized the utility function. Other models present more difficulty for language acquisition. That in itself should not be interpreted to mean that they are wrong as each has its own benefits.

The tied constraints model has the benefit of reducing the final grammar almost as much as acquiring constraints reduces it. However, if a tied constraints model is accurate, then the values of the coefficients in the model must be acquired as well. Because of the possibility of constraint tying, the coefficients are no longer regular.

Multiple constraint rankings likewise have additional features that need to be acquired, such as the maximal number of violations. This occurs regardless of whether the speaker is merely tallying, or actually calculating the number that is useful. This increases the numerical size of the grammar but reduces the number of variables that need to be manipulated. As we've seen, trading off features of UG and additional complexity in acquisition may lead to models of grammar that are ultimately easier to manipulate once learned.

The stochastic model presents the greatest challenge for learning. I assume here that the p_j values for the probability of a constraint varying begin with a value of zero. Before the

variation can be considered the constraint ranking must be established. If we assume that irregularities are established after regular behaviours, then it is clear that once the constraint ranking is established, the p_j values can be adjusted where needed to account for nuances. I assume for the moment that the probabilities would be adjusted via Bayesian principles, and if they are established only after the constraint ranking, it is reasonable to predict that this portion of the grammar may be adjustable over time, even while the constraint ranking itself remains fixed.

An alternative approach to the stochastic model is that the stochastic portion is the source of probabilistic behaviour, and that these probabilities diverge from zero very early, only to have the constraint ranking imposed upon a purely probabilistic model at a later date. More research into language acquisition will have to be done to determine which of these is a more accurate model of learning behaviour. Without the constraint ranking, however, the grammar is no longer Optimality Theoretic.

3.3. Multiple constraint rankings

As we mentioned in the discussion of the model of strict OT, multiple constraint rankings are possible for a given language. The theory tells us that constraints must be ranked, but that constraints that don't interact in a given language may be ranked in one order in one speaker's grammar, but ranked in a slightly different order in another's. A model that permits tied constraints, as described in this paper, does not require non-interacting constraints to be tied. Such a requirement would help reduce the size of the grammar and reduce or eliminate differences in the grammars across speakers of a single language. More than these minor variations, however, it may be also be possible to produce identical linguistic outputs but appealing to very different constraint interactions (McCall, 2002). These models do not make any predictions about how this might occur or how the utility values may differ. However, it should be possible to test in each case how accurate the predictions of each model are by conducting experimental studies in the field, and modeling the behaviour of each model to determine which values for p_j work best, and which models match the study's behaviour most closely. If it can be shown by these or other means that multiple rankings exist, the notion of language change through constraint reranking becomes, at the least, more complex than currently envisioned.

3.4. Linearity and nonlinearity in OT

The mathematical models described here also suggest another feature of Optimality Theory, which is a strong linear quality. While there is some allowance for constraint conjunction, the variables for the conjoined constraints are also zero or one. Gradience effects, while linear in individual constraints are the first suggestion of possible nonlinearity in Optimality Theory when we begin to consider conjoining them. However, nonlinearities are concealed in OT in the guise of output-output faithfulness constraints and sympathy theory. Sympathy theory, in particular, is language specific, and amounts to a clever way of masking constraint interaction. As we have seen from the discussion of gradience constraints that gradience, as difficult as it is for Optimality Theory, comes with certain advantages, one of these being to reduce the overall number of possible constraints. Likewise, by permitting more complicated interactions among constraints, further reductions may be possible, at the cost of additional complexity in the model.

3.5. Dynamics

Optimality Theory postulates two functions, EVAL and GEN. Most of this paper has been dedicated to discussing the EVAL function. However, the analysis of the EVAL function may bear directly upon an analysis of the GEN function in OT. GEN is the function which generates candidates for EVAL to evaluate, and it is usually seen as generating an infinite number of candidates which EVAL considers in parallel. However, a human brain cannot, in fact, evaluate an infinite number of candidates simultaneously. This is another problem that has been referred to as the problem of the infinities. A mathematical model of EVAL predicts that there will be some minimal utility value that can be a winning candidate. By allowing the two functions to interact, we can make GEN more efficient, and more difficult to modify once the grammar has been established. By preventing GEN from providing candidates to EVAL that have no chance of succeeding, such a model may provide another explanation for why second language acquisition is so difficult, since GEN may not be capable of even supplying winning candidates.

The stochastic model also has something to say about language dynamics over a speaker's lifetime, and for language change. If the value of p_j is adjusted in a Bayesian fashion over the life of the speaker, changes in the linguistic environment can be learned and the language of the speaker adjusted, even while the constraint ranking for that speaker remains fixed. Within a limited domain, new speakers might perceive the constraint ranking as already adjusted—even when variation still exists. We need only have some $p_j > 0.5$ to cause a change in the constraint ranking, since in language acquisition we assume that p_j would be adjusted upwards from zero.

4. Conclusions and future research

One can see that the mathematical models of Optimality Theory described here show in detail some of the theoretical consequences that variations on a basic theme can have. A strict model of OT has benefits that arise from its simplicity, but it forces grammars under the assumption of UG to be extremely large in relation to other models. The stochastic version of Optimality Theory shows the greatest promise for maintaining behavioural features of other models, and still being capable of adding new features to tackle linguistic variation across speakers, within a speaker's grammar over time and through the process of language change. Such mathematical models in general provide a concrete means of constraining aspects of the theory and using OT in other fields of language modeling such as natural language processing and producing simulations of studies to better determine whether the model proposed can actually produce the observed behaviours. Furthermore, the models help us see best where theoretical tools such as sympathy theory and other features introduce nonlinearities into a model that is otherwise very linear. This allows us to begin asking questions about these features if they do not point in a new direction for linguistic theory beyond OT.

References

- Antilla, A 1995 *Deriving Variation from Grammar: a study of Finnish genitives*. ROA-63
Archangeli D & Langendoen DT eds. 1997 *Optimality Theory: an Overview* (Blackwell: Malden, MA)
Berger JO 1985 *Statistical Decision Theory and Bayesian Analysis*, 2nd ed (Springer: New York)
Boersma P & Escudero P 2002 *Optimality-Theoretic modeling of microvariation in phonological perception and production*

Bonilha G 2002 Conjoined Constraints and Phonological Acquisition ROA-533-0802
Chernoff H & Moses LE 1959 Elementary Decision Theory (Dover Publications: New York)
Goldsmith JA 1995 The Handbook of Phonological Theory (Blackwell: Malden, MA)
Hammond M 2000 The Logic of Optimality Theory ROA-390-0400
Kleinbaum DG, Kupper LL, Muller KE & Nizam A 1998 Applied Regression Analysis and
Other Multivariable Methods (Duxbury Press: New York)
McCall B 2002 Solving Palatalization in Japanese Mimetics, ms.
McCarthy J 2002 Against Gradience ROA-510-0302
Prince A & Smolensky P 1993 Optimality Theory: Constraint Interaction in Generative
Grammar ROA-537-0802

Bi-directionality in OT

Form-Meaning Asymmetries and Bidirectional Optimization

David I. Beaver[†] and Hanjung Lee[‡]

Stanford University[†]

University of North Carolina at Chapel Hill[‡]

Abstract. This paper discusses architectural aspects of various versions of bidirectional OT so far proposed, and their treatments of blocking and other phenomena involving asymmetric relationships between form and meaning. The models to be studied here are the strong and weak bidirectional OT of Blutner (2000), and the asymmetric OT model of Wilson (2001). We show that each of these models provides at best a partial solution to the problems of form-meaning asymmetries. We argue that some of the problems of existing OT models can be eliminated using a variant of system which performs only one iteration of the Weak OT process.

1. Introduction

Bidirectional Optimality Theory allows us to see a wide range of problems which would previously have been considered unrelated from a new perspective, the perspective of asymmetric relationships between input and output. For interpretation, the input is a form and the output a meaning, and for production the input is a meaning and the output is a form. A mismatch is any case where there is no isomorphism between the space of meanings and the space of forms, say because one form has no meaning, or multiple meanings, or because a meaning is inexpressible, or may be expressed in multiple ways. From this perspective, we can understand the phenomenon of blocking as a process which prevents or removes form-meaning asymmetries.

In this paper, we study how various versions of OT handle mismatches, concentrating on the phenomenon of blocking. In section 2 we will be considering simpler, relatively standard OT architectures. The first two of these are unidirectional. What we will term *naïve OT production* is the approach seen in most OT syntax papers, and is close to the model that is used in OT phonology. Naïve OT production starts with some representation of meaning as input, and a set of candidate outputs provided by a function referred to as GEN. A set of linearly ranked constraints is then used to select between candidate surface forms. The second unidirectional approach, not surprisingly, works the other way: we will term it *naïve OT comprehension*, although Hendriks and de Hoop (2001) term it *OT semantics*. The input is a surface form, GEN offers a set of candidate meanings, and the linearly ranked constraint set is used to find the best meaning for the given form.

Some OT architectures provide grammars that cannot be reduced to a set of meaning-form pairs. One of these, which we will term *naïve back-and-forth OT*, consists of an obvious combination of naïve OT production and comprehension: the first is used for production only, and the second for comprehension only, an architecture discussed by Hendriks and de Hoop (2001). Note that even if the constraints used in each direction are the same, this model may not assign a consistent relation between meanings and forms. In particular for some choices of constraints, if you take a meaning, apply naïve OT production to get a form, and then apply naïve OT comprehension, you may not get back to the original meaning.

In addition to the three *naive* models, we will also consider four more sophisticated variants, sophisticated in the sense that they have been specifically designed to target some of the mismatch phenomena we will be discussing. The four other models to be studied are the *strong bidirectional* OT and *weak bidirectional* OT of Blutner (2000), the *asymmetric* OT model of Wilson (2001) and a model we will term *medium strength OT*, developed by Beaver (to appear). We will introduce these models individually later in the paper.

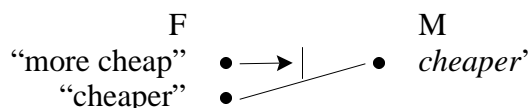
2. Blocking in Unidirectional Optimization Models

In this section we will consider blocking, which is the focus of this paper, and discuss the significance of this phenomenon for naive OT architectures.[‡]

Total Blocking

One of the classic cases of blocking is where the existence of a lexical form produced by productive morphology blocks a phrasal form. For instance, consider English comparative and superlative adjectival inflections: the existence of “cheaper” can be said to block “more cheap”, whereas the absence of “expensiver” means that “more expensive” is available (Poser 1992; Bresnan 2001):

- (1) a. cheaper/cheapest, ?more/?most cheap
b. *expensiver/*expensivest, more/most expensive



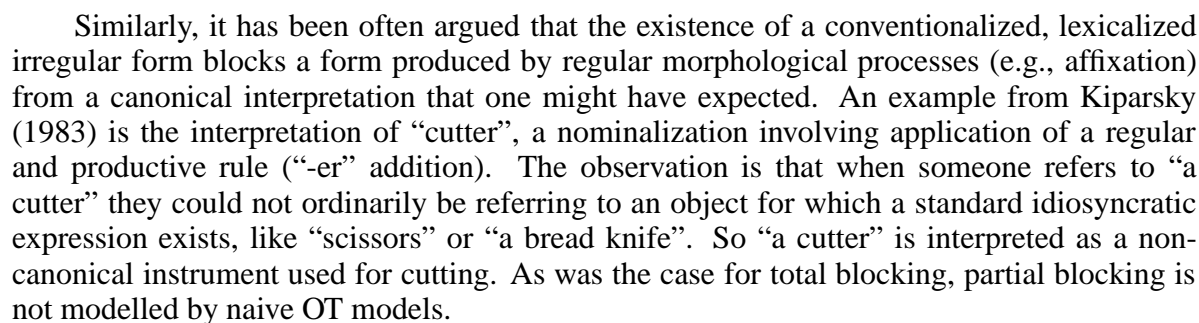
We can also understand cases involving alternative binding possibilities for pro-forms in terms of blocking of meaning (Levinson 2000). In Marathi, for example, a preference for more local anaphora resolution prevents resolution outside of the clause, as in (2a); resolution outside the clause is possible only when there is no blocking, as in (2b) (Dalrymple 1993:19–20):

- (2) a. Tom_i mhanat hota [ki Sue_j ni swataahlaa_{*i/j} maarle]. [Marathi]
Tom said that Sue ERG ANAPHOR-ACC hit
‘Tom said that Sue hit herself/*him.’
b. Jane_i mhanaali [ki [swataaci_i pariksha] sampli].
Jane said that ANAPHOR-GEN test finished
‘Jane said that her test was over.’

The existence of blocked meanings is not modelled by naive production OT, since it makes no prediction about which interpretation of the same form should be preferred. Similarly, blocking of forms is not predicted, if we take the interpretation perspective alone.

[‡] The discussion in the present paper closely follows the exposition given in Beaver and Lee (to appear), in which more extensive reviews of OT models are presented.

Blocking can leave a form unemployed, but the unemployed form may soon find a new job, generally expressing something closely related to but subtly different from the canonical interpretation that one might have expected. This is partial blocking: an asymmetry is eliminated, but removal of a link creates a new form-meaning pair. An example from McCawley (1978) is that of causatives. The observation is that the existence of a lexical causative “kill” blocks “cause to die” from having its canonical meaning. “Cause to die” comes to denote a non-canonical killing, for instance one where the chain of causation is unusually long or unforeseeable.



3. Strong Bidirectional Optimization

Strong OT offers a treatment of total blocking. Suppose that we are analyzing two forms f_1 and f_2 which are semantically equivalent and that we have some meaning m_1 that is optimal for both forms. In interpretation optimization, the two forms would not belong to the same candidate set and thus would both be grammatical. In the Strong OT model, f_2 , even if optimal

§ For applications of bidirectional OT to other cases of form-meaning asymmetries, see Smolensky (1998), Zeevat (2000), Asudeh (2001), Lee (2001), Vogel (to appear), among others.

in the interpretation-based optimization, may be blocked by the more economical alternative form f_1 . Hence, the form-meaning pair $\langle f_2, m_1 \rangle$ is removed from the set of the language generated by the Strong OT system.

Strong OT also opens up a simple way of modeling blocking of meaning. Consider the Marathi example in (2a) above. This sentence has the form $[A_i \dots [\delta B_j \dots \text{anaphor} \dots]]$, in which A and B are potential antecedents for the anaphor and δ is the domain in which the anaphor must have an antecedent (the minimal finite clause that contains the anaphor). Parsing this sentence will result in two classes of analyses: one in which the binding relation is local (i.e., anaphor = j) and one in which the binding relation is non-local (i.e., anaphor = i). In production-based optimization, the two interpretations do not compete with each other and thus the sentence is grammatical for both interpretations. In interpretation-based optimization, the former interpretation is preferred to the latter interpretation by a locality constraint on binding. As a result, anaphora resolution outside the clause is blocked by local anaphora resolution and hence removed from the set of interpretations generated by the Strong OT system. Taking together the two directions of optimization, we correctly predict not only that (2) is interpreted as $\text{say}(\text{Tom}, \text{hit}(\text{Sue}, \text{Sue}))$, but that it is the preferred way of expressing this meaning.

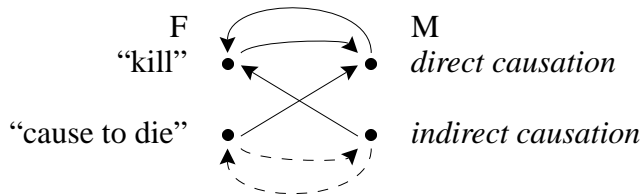
However, Strong OT fails to predict partial blocking. For example, strong OT predicts that “cause to die”, since it is blocked by the lexicalized “kill”, should be uninterpretable. But in fact it is only partially blocked, and comes to have an application in situations where “kill” would be deemed inappropriate. We now turn to Blutner’s proposed solution to this problem.

4. Weak Bidirectional Optimization

Blutner’s *weak* notion of optimality, which we refer to simply as Weak OT, is an iterated variant of Strong OT that produces partial blocking instead of strict blocking. In Weak OT, sub-optimal candidates in a strong bidirectional competition can become winners in a second or later round of optimization.

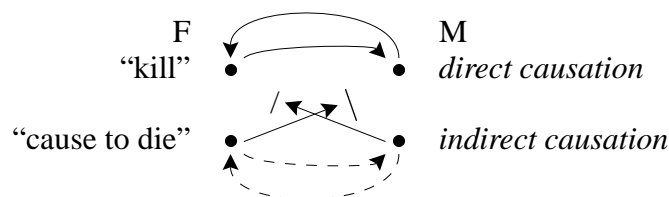
We illustrate how Weak OT predicts partial blocking using the example of lexical and periphrastic causatives “kill”/“cause to die” which we assume are matched on the meaning side by two possible interpretations, direct causation (canonical killing) and indirect causation (non-canonical killing). The following three diagrams, illustrate three phases of weak optimization. In the first diagram, all the unidirectionally optimal links are shown. In addition to the optimal links, two links are shown with dashed lines. Both of these links are unidirectionally sub-optimal at this stage, beaten by other candidates.

PHASE 1 — NAIVE INTERPRETATION AND PRODUCTION:



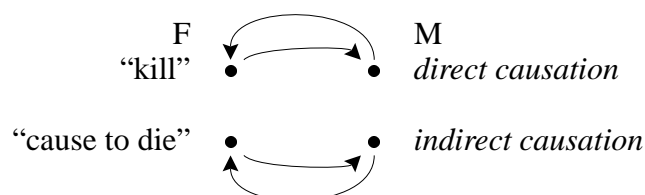
In phase 2 of Weak optimization, two unidirectionally optimal links are blocked, leaving a single bidirectionally optimal link, that between the form “kill” and the meaning corresponding to direct causation.

PHASE 2 — PRUNING:



Now we graft the originally sub-optimal links between “cause to die” and the indirect causation meaning back into the picture, since the candidates which originally beat them have been removed by blocking. This gives us two bidirectionally optimal links. In the resulting happy picture, all the candidate meanings are uniquely expressible and all the candidate forms are uniquely interpretable:

PHASE 3 — GRAFTING:



Blutner (2000) argues that Weak OT captures the essence of the pragmatic generalization that “unmarked forms tend to be used for unmarked situations and marked forms for marked situations” (Horn 1984:26). As Beaver and Lee (to appear) point out, however, Weak OT suffers from a serious problem of over-generation. Specifically, the process of adding extra links will eventually provide links for every form (if there are at least as many forms as meanings), or every meaning (if there are at least as many meanings as forms).

The problem of over-generation just mentioned obviously affects accounts of other phenomena involving form-meaning asymmetries. First, note that Weak OT fails to predict total blocking. While in the first phase of optimization the successful Strong OT predictions appear to be reproduced, in latter stages peculiar new form-meaning pairs will emerge as winners. Provided the set of candidate meanings is large, Weak OT never predicts total blocking: all blocking is partial. So a form like “more cheap”, for example, would presumably be the correct expression of some meaning in Strong OT.

Furthermore, Weak OT does not predict the existence of ineffable meanings and uninterpretable forms. For example, in Italian, multiple wh-questions are infelicitous for most speakers (Legendre, Smolensky and Wilson (1998)). Yet in this case Weak OT predicts that a multiple question is expressible since the grafting stage of Weak OT can add links to make it expressible. Uninterpretability is not predicted either since an uninterpretable form can be linked to a meaning by the grafting process.

5. Asymmetric Bidirectional Optimization

Wilson (2001) discusses a model in which interpretation precedes production. We refer to this as Asymmetric OT. (For discussion of different asymmetric models, see Zeevat (2000) and Vogel (to appear).) In more detail, the idea of Asymmetric OT is as follows: (i) Interpretation: Given any form-meaning pair $\langle f, m \rangle$, find the most harmonic semantic interpretation of f . (ii) Production: Given input meaning m , take as candidate outputs the set of forms f such that $\langle f, m \rangle$ is optimal in stage one, and perform standard OT production optimization with this restricted candidate set. Note that the set of optimal form-meaning pairs in production is

a subset of the optimal form-meaning pairs in interpretation. The set of meanings which are in some optimal pair is the same in interpretation and production, although the number of forms would, for constraint sets which are of interest, be smaller in production than in comprehension. It is the reduced set of forms in production, those which result from the two stage process, which are to be considered grammatical, even though there are others which are interpretable.

Wilson (2001) uses this version of OT to model partial blocking involving relativized minimality (see the examples in (2)) and referential economy in anaphor binding. An example of a referential economy effect is provided by the following contrast between the Icelandic third-person pronoun *hann* and the anaphor *sig*:

(3) Referential economy in Icelandic (Maling 1984: 212)

- a. Haraldur_i skipaði mér að raka *hann_i/sig_i.
Harold ordered me to shave him/ANAPHOR
'Harold ordered me to shave him.'
- b. Jón_i veit að María elskar hann_i/*sig_i.
Jon knows that Maria loves him/ANAPHOR
'Jon knows that Maria loves him.'

In (3a), the matrix subject *Haraldur* can grammatically bind the anaphor but not the pronoun. In (3b), in contrast, the pronoun is grammatical. According to Wilson, contrasts like the one in (3) follow from an interaction of two constraints: the LOC(AL) ANT(ECEDENT) constraint, which is a locality requirement on anaphor binding, and the REF(ERENTIAL) ECON(OMY) constraint, which requires a bound element to be an anaphor.

For the anaphora data above, the consequence of Asymmetric OT is as follows: for the interpretation optimization based on the string containing an anaphor, REFECON has no effect, since all candidates contain a bound anaphor. Thus, LOCANT gives us a local binding interpretation as the winner. In the interpretation optimization with the string containing a pronoun as the input, both local and nonlocal binding interpretations have the same constraint profile for REFECON and LOCANT, so both are selected as winners. The production optimization which takes nonlocal binding as input (Tableau 1), however, does not include the form containing an anaphor in the candidate set, since nonlocal binding loses in the interpretation competition with this form as input. As a result, the candidate with a pronoun wins trivially, and the more marked meaning, i.e., nonlocal binding, is predicted to be realized as a more marked (less economical) form. Note that the production tableau for local binding interpretation (Tableau 2) contains both forms, so this meaning is still realized as a form containing an anaphor:

Tableau 1. Production I (Asymmetric OT)

	REFECON	LOCANT
Input: nonlocal binding (m_2)		
☞ b. $[A_i[\delta B_j \dots \text{pronoun}_i]] (\langle f_2, m_2 \rangle)$	*	

Tableau 2. Production II (Asymmetric OT)

		REFECON	LOCANT
	Input: local binding (m_1)		
☞	a. $[A_i[\delta B_j \dots \text{anaphor}_j]] (\langle f_1, m_1 \rangle)$		
	b. $[A_i[\delta B_j \dots \text{pronoun}_i]] (\langle f_2, m_1 \rangle)$	*	

So far we have looked at the Asymmetric OT analysis of partial blocking in anaphor binding. Asymmetric OT, however, fails to model the standard cases of partial blocking discussed earlier. What distinguishes Wilson’s anaphora data is that the pair of a marked form and an unmarked meaning ($\langle f_2, m_1 \rangle$ in the above tableaux) and the pair of a marked form and a marked meaning ($\langle f_2, m_2 \rangle$ in the above tableaux) have the same constraint profile for the constraint favoring a less marked meaning (see the tableaux above). As noted above, the LOCANT constraint, preferring local binding over nonlocal binding, targets only an anaphor (f_1) but not a pronoun (f_2). As a result, the pairs $\langle f_2, m_1 \rangle$ and $\langle f_2, m_2 \rangle$ both survive in interpretation. Now when we come to realize m_1 , we don’t choose f_2 but instead choose f_1 . In other words, in production, as illustrated in the tableaux above, the pair $\langle f_1, m_1 \rangle$ blocks $\langle f_2, m_1 \rangle$, making $\langle f_2, m_2 \rangle$ available.

The standard cases of partial blocking differ in that the two pairs $\langle \text{marked form}, \text{unmarked meaning} \rangle$ and $\langle \text{marked form}, \text{marked meaning} \rangle$ do not have the same constraint profile (In Tableau 3, ECONOMY is a formal markedness constraint (a preference for short forms), and CANON is a semantic markedness constraint (a preference for the canonical mode of causation)):

Tableau 3. Interpretation (Asymmetric OT)

		ECONOMY	CANON
	Input: <i>cause to die</i>		
☞	a. $\langle \text{cause to die}, \text{direct causation} \rangle$	*	
	b. $\langle \text{cause to die}, \text{indirect causation} \rangle$	*	*

Asymmetric OT, while successfully modelling total blocking and certain cases of partial blocking that are interpretation-driven, fails to predict the full “division of pragmatic labor” whereby more marked forms are associated with more marked meanings. The constraints above yield a preferred interpretation of “cause to die” as involving canonical direct causation. Therefore, in the production competition with indirectly caused death as input meaning, “cause to die” is not even amongst the candidate outputs, and cannot be the winner.

We can see the difference between the two cases, and how they are treated, graphically. Diagrams (i–v), below, show both production and interpretation relations. The first two diagrams represent direct applications of naive back-and-forth OT: the first illustrates standard partial blocking cases yielding marked meanings for marked forms such as “cutter” and “cause to die”. The second diagram represents the situation Wilson describes for Icelandic anaphora. The only difference is an extra arrow from the marked form to the marked meaning in the second diagram.

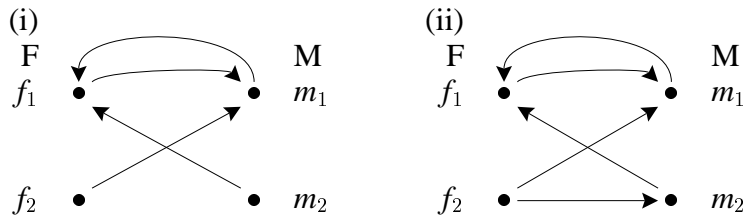
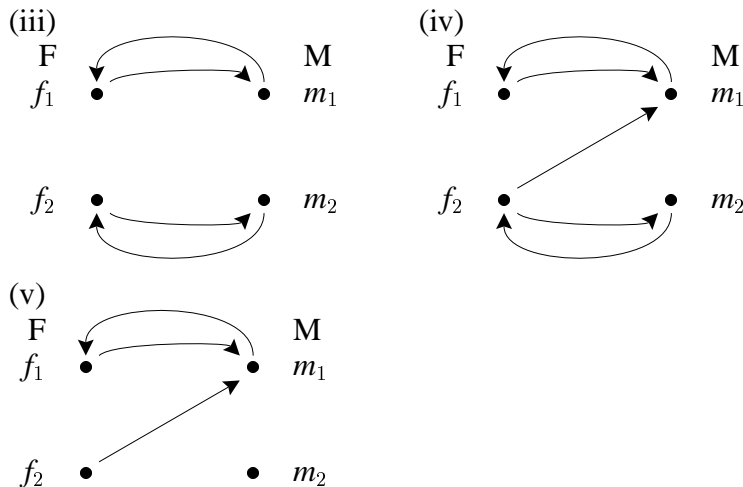


Diagram (iii) shows the results of applying Weak OT to either the situation in (i) or that in (ii): the marked form becomes uniquely associated with the marked meaning in both directions of optimization, while the unmarked form and unmarked meaning continue to be a bidirectionally optimal pair as they were in the original cases. Asymmetric OT does not achieve the harmonious situation depicted in (iii) for either of the situations given by (i) and (ii). What it does achieve is represented in (iv) and (v). Diagram (iv) shows the results of applying Asymmetric OT (IP) to the Icelandic anaphora case in (ii). Here we see that the division of labor depicted in (iii) is almost achieved, except that there remains the possibility of interpreting the marked form as the unmarked meaning. This is a result of the fact that Wilson's proposal does not innovate above naive back-and-forth OT as regards interpretation. When Asymmetric OT is applied to the classic "cause to die" situation in (i), what results is (v). Wilson's system does not succeed in creating any link between the marked form and the marked meaning, so we can see that it does not provide a very general model of partial blocking. In these cases we might better describe what it does as "almost blocking".



6. Medium Strength Optimization

It was noted above that Weak OT suffers from a serious problem of over-generation, as well as providing a problematic solution to total blocking. Could a variant of Weak OT maintain the analysis of partial blocking without such great over-generation? The possibility we will consider here is the variant of Weak OT discussed by Beaver (to appear). This variant system, which we refer to as Medium Strength OT, performs only one iteration of the Weak OT process, pruning once and grafting once. As a result, it maintains some of the properties of Weak OT, but lacks Weak OT's "everyone's a winner" profligacy.

In more detail, Medium Strength OT operates as follows. (i) starting with a set of production links and a set of interpretation links, find strong bidirection optimal form-meaning pairs. (ii) mark form-meaning pairs that have identical form or meaning to a bidirectionally optimal pair, but worse constraint violations. (iii) recalculate production and interpretation

links for the remainder to get a new set of strong bidirection optimal pairs. The set of medium strength winners is just the union of the winning sets from each round.

Stage (ii) corresponds loosely to the pruning phase (phase 2) of Weak OT. In Medium Strength OT, the recoverability condition on optimality (Smolensky 1998) is implemented into the model as a meta-linguistic constraint that acts as a blocking mechanism in the pruning phase. Let us term this ***BLOCK**, defined as follows:

- (4) ***BLOCK**: A form-meaning pair may not be dominated by (i.e., loses out to) a bidirectionally optimal candidate in either direction of optimization in the tableau consisting of all constraints except ***BLOCK**.

We illustrate how Medium Strength OT predicts partial blocking using the example of the Icelandic anaphora discussed in section 5. Consider first the following bidirectional tableau, in which the ***BLOCK** column is blank, but other constraint violations are marked. Candidate (a), with a locally bound anaphor, emerges immediately as a bidirectionally optimal form-meaning pair:

Tableau 4. Partial blocking in Medium OT I

		*BLOCK	REFECON	LOCANT
✎	a. $[A_i[\delta B_j \dots \text{anaphor}_j]] \langle f_1, m_1 \rangle$			
	b. $[A_i[\delta B_j \dots \text{pronoun}_j]] \langle f_2, m_1 \rangle$		*	
	c. $[A_i[\delta B_j \dots \text{anaphor}_i]] \langle f_1, m_2 \rangle$			*
	d. $[A_i[\delta B_j \dots \text{pronoun}_i]] \langle f_2, m_2 \rangle$		*	

Now let us consider how violations of ***BLOCK** are evaluated. Of the three candidates that are originally non-optimal, candidates (b) and (c) have identical form or meaning to the bidirectionally optimal candidate (candidate (a)), but worse violations of the standard constraints. Hence they are marked with a star in the ***BLOCK** column, as shown in Tableau 5:

Tableau 5. Partial blocking in Medium OT II

		*BLOCK	REFECON	LOCANT
✎	a. $[A_i[\delta B_j \dots \text{anaphor}_j]] \langle f_1, m_1 \rangle$			
	b. $[A_i[\delta B_j \dots \text{pronoun}_j]] \langle f_2, m_1 \rangle$	*	*	
	c. $[A_i[\delta B_j \dots \text{anaphor}_i]] \langle f_1, m_2 \rangle$	*		*
✎	d. $[A_i[\delta B_j \dots \text{pronoun}_i]] \langle f_2, m_2 \rangle$		*	

Thus Medium Strength OT produces two bidirectionally optimal candidates, $\langle f_1, m_1 \rangle$ and $\langle f_2, m_2 \rangle$, so we can see that it successfully predicts the full ‘division of pragmatic labor’ whereby more marked forms are associated with more marked meanings. The same result occurs with the standard cases of partial blocking, so no tableau will be shown here.

Although we will not provide detailed analyses here, it should be obvious that Medium Strength OT can model ineffability and uninterpretability: the one extra round of optimization produces some new pairs, but it does not produce anything as weird as “colorless green ideas” or “floodlsnoop”, and it need not produce a short way of expressing multiple questions like “Who ate what?” in Italian.

7. Conclusion

Most previous bidirectional OT models have failed to model the full range of blocking phenomena. The one system which does model the full range, Blutner’s Weak system, does so only at the expense of massive over-generation, making it untenable as a model of online interpretation or production. The Medium Strength system is a compromise between Weak and Strong OT. The compromise can be understood in terms of the following restatement of these three notions of optimality:

Strong The set S of strongly optimal form-meaning pairs is the largest set (of form-meaning pairs) which are undominated in interpretation and undominated in production.

Weak The set W of weakly optimal form-meaning pairs is the largest set which is undominated by other weakly optimal form-meaning pairs in interpretation and undominated by other weakly optimal form-meaning pairs in production.

Medium The set M of medium-strength optimal form-meaning pairs is the largest set which is undominated by other strongly optimal form-meaning pairs in interpretation and undominated by other strongly optimal form-meaning pairs in production.

By these definitions it is clear that $S \subseteq M \subseteq W$. Strong OT, like Asymmetric OT, does not produce enough form-meaning pairs to account adequately for partial blocking. Weak OT produces enough for partial blocking, but also produces many form-meaning pairs which have no linguistic significance. So the question is, does Medium Strength OT yield enough pairs, and does it yield too many pairs. This is an empirical question.

Suppose that form-meaning pairs created as a result of partial blocking were known synchronically to cause yet further partial blocking. A hypothetical case would be if use of “cause X to die” to refer to indirectly caused death prevented “lead to the death of X ” from having this meaning, and caused the latter locution to have yet another interpretation. Such a chain of partial blocking would constitute a counterexample to Medium Strength OT and force us to move further along the hierarchy towards Weak Bidirectional OT. However, we are not currently aware of any attested counter-examples of this sort. Thus we offer Medium Strength OT as a working hypothesis as to how interpretation and production interact to co-determine what is optimal in human language.

References

- Asudeh, Ash. 2001. Linking, optionality and ambiguity in Marathi: An Optimality Theory analysis. In Peter Sells (ed.), *Formal and Empirical Issues in Optimality Theoretic Syntax*, 257–312. Stanford: CSLI Publications.
- Beaver, David. to appear. The optimization of discourse anaphora. *Linguistics and Philosophy*.
- Beaver, David and Hanjung Lee. to appear. Input-output mismatches in OT. In Reinhard Blutner and Henk Zeevat (eds.), *Optimality Theory and Pragmatics*. Palgrave/Macmillan.

- Blutner, Reinhard. 2000. Some aspects of optimality in natural language interpretation. *Journal of Semantics* 17: 189–216.
- Bresnan, Joan. 2001. Explaining morphosyntactic competition. In Mark Baltin and Chris Collins (eds.), *Handbook of Contemporary Syntactic Theory*, 11–44. Oxford: Blackwell.
- Dalrymple, Mary. 1993. *The Syntax of Anaphoric Binding*. Stanford: CSLI Publications.
- Hendriks, Petra and Helen de Hoop. 2001. Optimality Theoretic semantics. *Linguistics and Philosophy* 24: 1–32.
- Horn, Lawrence. 1984. Toward a new taxonomy of pragmatic inference: Q-based and R-based implicature. In D. Schffrin (ed.), *Meaning, Form and Use in Context*, 11–42. Washington, DC: Georgetown University Press.
- Jäger, Gerhard. to appear. Learning constraint sub-hierarchies. The bidirectional gradual learning algorithm. In Reinhard Blutner and Henk Zeevat (eds.), *Optimality Theory and Pragmatics*. Palgrave/Macmillan.
- Kiparsky, Paul. 1983. Word-formation and the lexicon. In F. Ingeman (ed.), *Proceedings of the 1982 Mid-America Conference*, 113–137.
- Kuhn, Jonas. 2001. *Formal and Computational Aspects of Optimality-theoretic Syntax*. Doctoral Dissertation, Institut für maschinelle Sprachverarbeitung, Universität Stuttgart.
- Lee, Hanjung. 2001. *Optimization in Argument Expression and Interpretation: A Unified Approach*. Doctoral Dissertation, Stanford University.
- Legendre, Géraldine, Paul Smolensky and Colin Wilson. 1998. When Is Less More? Faithfulness and Minimal Links in *wh*-Chains. In P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis and D. Pesetsky (eds.), *Is the Best Good Enough? Optimality and Competition in Syntax*, 249–289. Cambridge: MIT Press.
- Levinson, Stephen. 2000. *Presumptive Meanings: The Theory of Generalized Conventionalized Implicature*. Cambridge: MIT Press.
- McCawley, James. 1978. Conversational implicature and the lexicon. In P. Cole (ed.), *Syntax and Semantics 9: Pragmatics*, 245–259. New York: Academic Press.
- Poser, William. 1992. Blocking of phrasal constructions by lexical items. In Ivan Sag and Anna Szabolcsi (eds.), *Lexical Matters*, 111–130. Stanford: CSLI Publications.
- Smolensky, Paul 1996. On the Comprehension/Production Dilemma in Child Language. *Linguistic Inquiry* 27: 720–731.
- Smolensky, Paul. 1998. Why Syntax is Different (But Not Really): Ineffability, Violability and Recoverability in Syntax and Phonology. Handout of the talk given at the Stanford/CSLI Workshop: Is Syntax Different? Common Cognitive Structures for Syntax and Phonology in Optimality Theory. Stanford University, December 12–13, 1998.
- Vogel, Ralf. to appear. Remarks on the architecture of OT syntax grammars. In Reinhard Blutner and Henk Zeevat (eds.), *Optimality Theory and Pragmatics*. Palgrave/Macmillan.
- Wilson, Colin. 2001. Bidirectional Optimization and the Theory of Anaphora. In Géraldine Legendre, Jane Grimshaw and Sten Vikner (eds.), *Optimality-theoretic Syntax*, 465–507. Cambridge: MIT Press.
- Zeevat, Henk. 2000. The Asymmetry of Optimality Theoretic Syntax and Semantics. *Journal of Semantics* 17: 243–262.

Partial Blocking, Associative Learning, and the Principle of Weak Optimality

Anton Benz

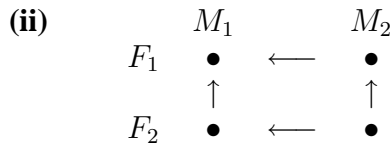
Zentrum für Allgemeine Sprachwissenschaft, Berlin, Germany

1. Introduction

One of the selling-points of Bi-OT is its success in explaining partial blocking phenomena. In (i) it has to explain why *kill* tends to denote a *direct* killing whereas *caused to die* tends to denote an *indirect* killing [6]:

- (i) a) Black Bart killed the sheriff.
- b) Black Bart caused the sheriff to die.

The Bi-OT explanation is based on the principle of weak optimality, a generalisation of a rule known as *Horn's division of pragmatic labour* [10, p. 22]: Marked forms typically get a marked interpretation, and unmarked forms an unmarked interpretation. *Kill* is the less marked form, and if we assume that speakers prefer less marked forms over marked forms, then *kill* is the optimal way to denote a killing event. As direct killing is the normal and expected way of killing, the hearer should have a preference for interpreting the speaker's utterance as referring to a direct killing. We can see that *kill* and *direct killing* build an optimal form-meaning pair from both perspectives. In addition we can see that the marked form tends to denote the less expected meaning, i.e. *cause to die* tends to denote an *indirect killing*. In general, if F_1 and F_2 are forms and M_1 and M_2 are meanings where F_1 is preferred over F_2 and M_1 over M_2 , then F_1 tends to denote M_1 and F_2 tends to denote M_2 :



Horn explains his principle by recursion to two pragmatic principles, called the Q- and R-principle. Blutner [5] gave them a formally precise formulation. Specifically, he made explicit the role of switching between speaker's and hearer's perspective. This laid the foundation for an optimality-theoretic reformulation, and thereby for placing radical pragmatics in the broader linguistic context provided by OT. In this paper we are going to explain partial blocking as the result of diachronic processes based on what we will call *associative learning*.

(1) Bi-OT over-generates partial blocking, i.e. it predicts partial blocking for many examples where blocking is not observable; (2) Bi-OT in its original form has only a weak foundation, i.e. there is no good explanation for the principle of weak optimality which does (a) not make an (implicit) appeal to Horn's principle of pragmatic labour, and (b) provides more than just an algorithm for how to calculate weakly optimal form-meaning pairs. Game theory has been proposed as a remedy for the last problem [9]. We will discuss Bi-OT at more length in Section 2, and in Section 3 we consider van Rooy's game-theoretic approach to explaining Horn's division of pragmatic labour [16]. Partial blocking can be observed in examples where expressions are unambiguous and where there would be an alternative form

for denoting the more marked meaning. We will see that these assumptions about language make van Rooy's model inapplicable.

Originally, Blutner understood his theory from a diachronic perspective[‡]. We take this idea more seriously. We claim that partial blocking can be explained as an effect of *associative learning* plus speaker's preferences on forms. It emerges as a result of a diachronic process. We explain Example (i) by postulating the following five stages: (1) In the initial stage all killing events are direct killing events. The speaker will always use *kill* to denote these events. (2) Interpreters will learn that *kill* is always connected with direct killing. They *associate kill* with direct killing. (3) The speaker will learn that hearers associate *kill* with direct killing. (4) If then an exceptional event occurs where the killing is an indirect killing, the speaker has to avoid misleading associations, and use a different form. In this case it is the more complex form *cause to die*. (5) The hearer will then learn that *cause to die* is always connected to an untypical killing. By *associative learning* we mean the learning process in (2), (3), and (5). We postulate the following principle related to the hearer:

In every actual instance where the form F is used for classifying events or objects it turns out that the classified event or object is at least of type t , then the hearer learns to associate F with t , i.e. he learns to interpret F as t .

A similar principle is assumed for the speaker to explain step (3). Given a set of semantically synonymous expressions, how can associative learning and speaker's preferences lead to a change in interpretation? In Section 4 we work out a formal model which describes diachronic processes related to associative learning.

2. Bi-OT and Weak Optimality

According to OT, producer and interpreter of language use a number of constraints which govern their choice of forms and meanings. These constraints may get conflict with each other. OT proposes a mechanism for how these conflicts are resolved. It assumes that the constraints are ranked in a linear order. If they get into conflict, then the higher-ranked constraints win over the lower-ranked ones. This defines preferences on forms and meanings.

Optimality theory has divided into many sub-theories and variations. Beaver and Lee [2] provide for a useful overview of versions of optimality-theoretic semantics. They discuss seven different approaches. In particular they compare them according to whether they can explain partial blocking. It turns out that the only approach which can fully justify Horn's division of pragmatic labour is Blutner's Bi-OT [2, Sec. 7 and 5].

What are the structures underlying Bi-OT? In bidirectional OT it is common to assume that there is a set \mathcal{F} of *forms* and a set \mathcal{M} of *meanings* [6]. A set Gen , the so-called *generator*, tells us which form-meaning pairs are grammatical. The grammar may leave the form-meaning relation highly underspecified. In a graphical representation like (ii) a grammatical form-meaning pair $\langle F, M \rangle$ is represented by a bullet at the point where the row for F and the column for M intersect. Underspecification means that a row corresponding to a form F may contain several bullets. The speaker has to choose for his utterance a form which subsequently must be interpreted by the hearer. It is further assumed that the speaker has some ranking on his set of forms, and the hearer on the set of meanings. Blutner [6] introduced the idea that the speaker and interpreter coordinate on form-meaning pairs which are most preferred from both perspectives. The speaker has to choose for a given meaning M_0 a form F_0 which is optimal according to his ranking of forms. Then the interpreter has to choose for F_0 a meaning M_1 which is optimal according to his ranking of meanings. Then again the speaker looks for the

[‡] Personal communication.

most preferred form F_1 for M_1 . A form–meaning pair is optimal if ultimately speaker and hearer choose the same forms and meanings. If $\langle F, M \rangle$ is optimal in this technical sense, then the choice of F is the optimal way to express M so that both speaker’s and interpreter’s preferences are matched.

It is easy to see that the procedure for finding an optimal form–meaning pair stops for a pair $\langle F, M \rangle$ exactly if there are no pairs $\langle F', M \rangle$ and $\langle F, M' \rangle$ such that the speaker prefers F' over F given M and the hearer prefers M' over M given F . In the graph (ii) $\langle F_1, M_1 \rangle$ is optimal because there are no arrows leading from $\langle F_1, M_1 \rangle$ to other form–meaning pairs. *Weak optimality* is a weakening of the notion of optimality. In (ii) we find that F_2 should go together with M_2 . For $\langle F_1, M_2 \rangle$ and $\langle F_2, M_1 \rangle$ there is either a row or a column which contains it together with the optimal form–meaning pair $\langle F_1, M_1 \rangle$. For $\langle F_2, M_2 \rangle$ neither its row nor its column contains the optimal $\langle F_1, M_1 \rangle$. If we remove the row and the column which contain $\langle F_1, M_1 \rangle$, then $\langle F_2, M_2 \rangle$ is optimal in the remaining graph. This can be generalised: If we remove from a given graph all rows and columns which contain an optimal form–meaning pair, then the optimal form–meaning pairs in the remaining graph are called *weakly optimal*. We can iterate this process until no more form–meaning pairs, and hence no graph, remains§.

The Problem of Over-Generation Bi-OT can successfully explain examples like (i) but if we apply it naively, then there are many examples where it over-predicts partial blocking. We first look at examples with anaphora resolution where it is semantically not clear who of the antecedents is male or female but where one of the alternatives is highly preferred. We don’t get a marked interpretation for a marked expression||:

- (iii) a) The doctor kissed the nurse. She is really beautiful.
b) The doctor kissed the nurse. The woman is really beautiful.
c) The doctor kissed the nurse. Marion is really beautiful.
d) (?)The doctor kissed the nurse. SHE is really beautiful.

If the hearer has no special knowledge about the doctor and the nurse, he will interpret the second sentence as meaning *the nurse is really beautiful*. If we assume further that a pronoun is more economic than a proper name, and a proper name more economic than a definite description, then the speaker should continue his first sentence with *She is really beautiful*. The uses of *Marion* and *the woman* are less preferred, hence they should go together with a marked interpretation. If we apply the principle of weak optimality straightforwardly, then it predicts a tendency of e.g. *Marion*, or *the woman*, to indicate that the doctor is a woman. But for all three examples we get the same reference. If we stress the pronoun, then the sentence becomes ungrammatical rather than getting a marked reading.

Examples (iii) and (iv) are cases where underspecification is crucially involved. The next two examples represent cases without underspecification:

§ The principle of weak optimality is due to Blutner, see [5, 6]. He calls *superoptimality* what was later called weak optimality. The process for finding weakly optimal form meaning pairs is due to G. Jäger, see [7, 11]. [9] was a first attempt to bring weak optimality together with the notion of *nash equilibria*.

|| Examples of this type have first been discussed by J. Mattausch [13].

- (iv) a) Hans hat sich ein Rad gekauft.
 b) Hans hat sich ein Fahrrad gekauft.
 c) Hans hat sich ein Zweirad gekauft.
 d) Hans has himself a bicycle bought.

The first two sentences are equivalent but the third is marked. The critical expressions are *Rad*, *Fahrrad* and *Zweirad*. In this context they have all the same meaning, namely *bicycle*. The principle of weak optimality would predict that *Rad* (wheel) is optimal, hence *Fahrrad* (driving-wheel) should tend to have a marked meaning. But both expressions are equivalent. *Fahrrad* and *Zweirad* (two-wheel) are of the same complexity, hence there should be no difference in meaning, but *Zweirad* is marked. In contrast, the following example clearly is in line with weak OT and Horn's principle of division of pragmatic labour:

- (v) a) Hans wischt den Boden mit Wasser/Flüssigkeit.
 b) Hans mops the floor with water/a liquid.

Flüssigkeit (liquid/fluid) clearly indicates that it is not water that Hans uses for mopping the floor.

We observe a difference between a class (A) with example (iv) where the hearer has to resolve an ambiguity for interpreting the speaker's utterance, and a class (B) where the critical expressions differ only with respect to their extension. Example (i) belongs to class (B), i.e. to examples (iv) and (v).

We have seen that we don't get the effects predicted by Bi-OT for class (A). Marked expressions don't show a tendency to go together with the unexpected reading. Our examples which show partial blocking belong all to class (B)[¶]. Conceptually, this is an important point as the assumption that meaning is highly underspecified is central for Bi-OT. Bi-OT in its naive form makes predictions for both classes.

3. Game Theory and Partial Blocking — van Rooy's Principle

We have seen in Section 2 that Bi-OT over-predicts partial blocking if applied too naively. Originally Blutner intended his theory not as a synchronic theory, i.e. as a theory which models the actual reasoning of interlocutors in an utterance situation. Weak optimality was intended to select diachronically stable form-meaning pairs. Soon after emergence of Bi-OT, Game Theory was proposed as a foundational framework [9]. It allows to embed OT within a well understood theory of rational decision. In addition, there has been important work by Prashant Parikh [14, 15] on resolving ambiguities within game theoretic frameworks. For the following discussion we concentrate on van Rooy's paper [16] because he explicitly proposes his theory as a game theoretic explanation of Horn's division of pragmatic labour. Our aim in this section is not so much to show weaknesses of this approach but to show that it applies to different problems.

For simplicity we represent the possible meanings as *attribute-value functions*; i.e. as functions $f : \text{Feat} \rightarrow \text{Val}$ from features into values $\{0, 1, -1\}$. Let m be some feature representing some property of objects, f an attribute-value function, and e an object of type f , $e : f$. Then $f(m) = 1$ means that e does have the property m ; $f(m) = -1$ means that e does not have the property m ; and $f(m) = 0$ means that e may or may not have the property m . We denote the set of all attribute-value functions by **Type**. $f \in \text{Type}^*$ means that all properties are specified. We call the elements of **Type**^{*} *basic types*. Attribute-value functions are very primitive examples of typed feature structures [8].

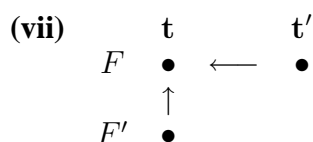
[¶] There is some work now on anaphora and OT starting with [1]. Examples of class (A) constitute a different type of problem. Hence we restrict our considerations to cases without ambiguities, i.e. class (B). I discussed Mattausch's examples in two previous papers, [3] and [4].

Semantics and pragmatics should tell us what are the optimal forms for the speaker to select and how the hearer interprets them. A *speaker's selection strategy* is a function from meanings into forms; and a *hearer's interpretation strategy* a function from forms into meanings.

Van Rooy observes that if communication shall be successful, i.e. if $H(S(t)) = t$, then speakers and hearers must coordinate on *separating* strategy pairs $\langle S, H \rangle$, i.e. there must be a subset of forms \mathcal{F}' such that $H \circ S$ maps \mathcal{F}' 1–1 onto \mathcal{M} . This implies that it is desirable that speaker's strategies are also *separating*, i.e. that $t \neq t'$ implies $S(t) \neq S(t')$. Only then can it be guaranteed that every state of affairs can be expressed by language. If the speaker's strategy is not separating, then communication must fail for at least one situation, i.e. there exists $t \in \mathcal{M}$ such that $H(S(t)) \neq t$. If it is rational for interlocutors to coordinate on strategies where communication is always successful, then the following principle must hold:

- (vi) Suppose that F is a lighter expression than F' , $F > F'$, and that F' can only mean t , but F can mean both. Suppose, moreover, that t is more salient, or more stereotypical, than t' , $t > t'^+$, then speaker and hearer coordinate on strategy pairs $\langle S, H \rangle$ such that $S(t) = F'$, $S(t') = F$, $H(F) = t'$ and $H(F') = t$.

Van Rooy introduces his principle as a counterexample for Bi-OT. We can represent the situation by the following graph:



It is not difficult to see that van Rooy's principle (vi) contradicts Bi-OT and Horn's division of pragmatic labour. Clearly $\langle F, t \rangle$ is optimal. If we then reduce the graph and eliminate all nodes in the row and column containing $\langle F, t \rangle$, then no combination remains. Hence, Bi-OT predicts that F denotes t — and t' cannot be expressed.

The following examples show that van Rooy's principle is violated in situations of class (B). The claim that interlocutors always coordinate on the separating strategy seems to be incorrect:

- (viii) a) Zwei Amerikaner wurden bei dem Anschlag getötet.
b) Mehrere Afrikaner wurden in der S-Bahn angepöbelt.
a) Two Americans were in the plot killed.
b) Some Africans were in the city train verbally abused.

Without special context these sentences must be understood as:

- a) Zwei US–Amerikaner wurden bei dem Anschlag getötet.
b) Mehrere Schwarzafrikaner wurden in der S-Bahn angepöbelt.
a) Two US Americans were in the plot killed.
b) Some Black Africans were in the city train verbally abused.

The critical expressions are *Amerikaner* and *Afrikaner*. They have a wider extension than *US–Amerikaner* and *Schwarzafrikaner*. Moreover, they are lighter than the special expressions and the special expressions can only have a special meaning. We can assume that (a) in most cases where Germans talk about inhabitants of the American continent, they talk about US Americans, and (b) Black Africans are more prototypical Africans than North

⁺ The first part is cited from [16, Sec. 3.2, p. 13]. The notation is slightly adapted.

Africans; furthermore we can assume that the difference between US-Americans and Non-US Americans and Black Africans and Non-Black Africans is relevant. If we naively apply van Rooy's principle, then we should expect a tendency for *Amerikaner* to denote Non-US Americans, for *Afrikaner* to denote North-Africans, etc. But we observe the opposite effect.

It is not confined to examples where we classify people according to their nationality:

- (ix) a) Hans macht Urlaub in *Amerika*.
 b) Hans fährt seinen *Wagen* in die Garage.
 c) Hans makes holidays in *America*.
 d) Hans drives his *car* into the garage.

The first example must be understood as meaning that Hans makes holidays in the USA, not e.g. in Chile. *Wagen* can have a very wide meaning including both a car and a hand cart. The lighter, more general expression has always the tendency to denote the normal case. What if van Rooy's principle could be applied to these examples? It would predict the contrary effect. Van Rooy's principle is violated in class (B) — if applied too naively, of course. By *applying naively* I mean: applying without checking the preconditions. There are two reasons for why van Rooy's models cannot be used for class (B). He has to assume that the meaning of some forms is underspecified. Then, he has to start with non-separating signalling systems, and try to show that they develop into separating ones. This implies that the models cannot be applied if:

- i. Forms have unique meanings.
- ii. Languages are separating.

This is the situation we find in examples of class (B). We can always assume that natural language is fine-grained enough to express every state of affairs, i.e. we can assume that natural language is separating. Hence, the central problem with partial blocking phenomena is to explain how there can be shifts in meaning for signalling systems that are (a) separating and (b) unambiguous. If this is true, then partial blocking poses a type of problem which is sharply differentiated from the problems approached by van Rooy or Parikh.

4. Associative Learning and Partial Blocking

For the introductory example (i) it has to be explained why *kill* tends to denote a *typical* killing event whereas *cause to die* tends to denote an *untypical* killing event. I want to show that partial blocking can be explained as an effect of *associative learning* and speaker's preferences. It emerges as the result of a process which divides into the following stages: (1) In the initial stage all killing events are direct killing events. The speaker will always use *kill* to denote these events. (2) Interpreters will learn that *kill* is always connected with direct killing. They *associate kill* with direct killing. (3) The speaker will learn that hearers associate *kill* with direct killing. (4) If then an exceptional event occurs where the killing is an indirect killing, the speaker has to avoid misleading associations and use a different form. In this case it is the more complex form *cause to die*. (5) The hearer will then learn that *cause to die* is always connected to an untypical killing. By *associative learning* we mean the learning process in (2), (3) and (5). For the hearer I assume that the following principle holds:

- (H) In every actual instance where the form *F* is used for classifying events or objects it turns out that the classified event or object is at least of type *t*, then the hearer learns to associate *F* with *t*, i.e. he learns to interpret *F* as *t*.

A similar principle is assumed for the speaker to explain step (3):

- (S) In every actual instance where the form F is used for classifying events or objects it turns out that the hearer interprets F as t , then the speaker learns that he can use F for expressing t .

It is not only word meaning that is involved:

- (x) The dress is pink/pale red/pale red but not pink.

All three phrases, *pink*, *pale red*, and *pale red but not pink*, are forms which the speaker can choose. The forms F may even be lengthy descriptions of a situation.

A formal model must contain the following elements: (1) A set of possible meanings for words and phrases. (2) A representation for the semantics of a given language NL . (3) A representation for the speaker's preferences on forms. We do this by adding a pre-order \preceq on NL , where $F \prec F'$ means that F is less marked than F' .

Less obvious from the previous discussion is that we will need also: (4) A representation for the speaker's knowledge about the object or event he wants to classify. (5) A representation for the speaker's intentions on how to classify an object or event.

We consider settings of the following form: There is an object or event e and the speaker wants to classify it as being of a certain type f' . Maybe he knows more about the object, maybe he knows that it is in fact of a more special type f . But all he wants to communicate is that it is of type f' . He has to choose a form F such that the hearer can conclude that the object or event e is of type f' . This explains why we need a representation for speaker's knowledge and intentions. We represent them by attribute-value functions.

These elements form the *static* part of our model. What does change diachronically? (6) The types of objects and events which actually occur. We represent the actual occurrences of objects and events during a period α by a set E_α . (7) The hearer's interpretation of forms. We represent it by a function H from forms into meanings. (8) The speaker's choice of forms. We represent it by a function $S : \langle f, f' \rangle \mapsto F \in NL$, i.e. a function which maps pairs of attribute-value functions which represent his knowledge (f) and intentions (f') into forms. We assume throughout that the speaker is truthful and sincere; this means especially that f' represents not more information than f . The functions S and H are the counterparts of the speaker's and hearer's strategies in game-theoretic approaches.

We noted in the last section that the central problem with partial blocking phenomena is to explain how there can be shifts in meaning for signalling systems that are (a) separating and (b) unambiguous. We assume that in the initial situation choice and interpretation of language is governed by its (unambiguous) semantics. Let us denote the meaning of a form F by $[F]$, and assume that for every meaning f there is at least one form F such that $[F] = f$. The speaker should select the optimal form:

$$S^0(f, f') := \min\{F \in NL \mid f \leq H^0(F) \leq f'\}.$$

The hearer's initial interpretation should simply follow the rules of pure semantics; i.e. $H^0(F) = [F]$. The definitions imply that

$$f \leq H^0(S^0(f, f')) \leq f', \quad (4.1)$$

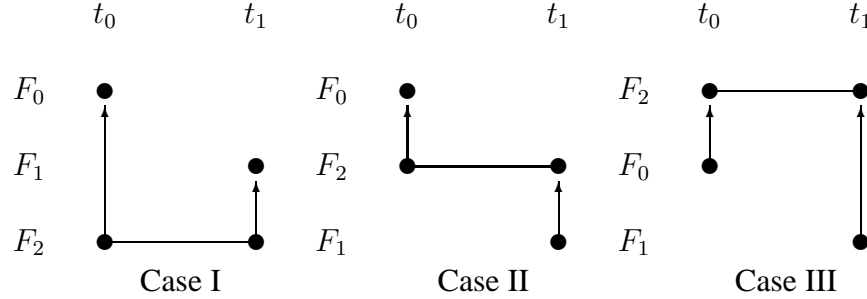
i.e. the speaker will always have success. In addition we assume that the speaker does classify entities correctly.

The Situation with two Basic Types

We look at a special case: the situation for one feature with two values. The examples considered so far are of this type, at least after some simplification of the scenarios. E.g. in (i) the question was whether the killing is *direct* or not. Hence we can assume one feature

direct with possible values -1 and 1 for *not direct* and *direct*. In (v) the question was whether it is *water* or not that Hans uses for mopping the floor.

If we consider a situation with two basic types t_0 and t_1 , then there are only three forms F_0, F_1, F_2 the speaker has to consider for making his choice. Without loss of generality we can assume that $[F_0] = t_0$, $[F_1] = t_1$ and $[F_2] = t_0 \vee t_1$. Hence, F_2 always denotes the form with the wider meaning. We can further assume that in general F_0 is preferred over F_1 . Hence, we arrive at the following classification of all situations with two basic types:



The topmost form is the most preferred one, the lowest the least preferred. The vertical arrow indicates the speaker's preferences. The horizontal line means that the respective form has an extension which comprises the meaning of both types t_0 and t_1 . Examples are: Case I *father, mother, one of the parents* ($F_0 \prec F_1 \prec F_2$); Case II *water, liquid, alcoholic essence* ($F_0 \prec F_2 \prec F_1$); Case III *American, North American, Latin American* ($F_2 \prec F_0 \prec F_1$).

Hence, we see that (v) is a Case II example. What about *kill-and-cause-to-die* (i)? We may assume that the relevant forms are $F_2 = \textit{killed}$, $F_0 = \textit{directly killed}$, and $F_1 = \textit{indirectly killed}$, hence it belongs to class III. For the classification we considered only the most economic forms for each type. We add $F_3 = \textit{caused to die}$ and assume for simplicity that $F_2 \prec F_3 \prec F_0 \prec F_1$. This is a sub-case of Case III. How can we explain the observed differentiations in meaning between F_2 and F_3 ? We claimed that we can see it as the result of a diachronic learning process. This process stretches over a sequence of (*synchronic*) *stages*. We have to describe how selection and interpretation strategies change from stage to stage. What is a synchronic *stage*? It is a triple $\text{Syn}_i = \langle E^i, S^i, H^i \rangle$ where

$$E^i \subseteq E \times \mathbf{Type} \times \mathbf{Type} \ \& \ \langle e, \mathbf{f}, \mathbf{f}' \rangle \in E^i \Rightarrow (e : \mathbf{f} \ \& \ \mathbf{f} \leq \mathbf{f}'). \quad (4.2)$$

This means that every synchronic stage is characterised by (1) the set of utterance situations which comprises a classified entity e , the speaker's knowledge \mathbf{f} about e , and his intentions to classify e as \mathbf{f}' ; (2) the speaker's selection strategy; and (3) the hearer's interpretation strategy.

We repeat the informal description of the principles governing the hearer's learning in each stage:

- (H) In every actual instance where the form F is used for classifying events or objects it turns out that the classified event or object is at least of type \mathbf{f} , then the hearer learns to associate F with \mathbf{f} , i.e. he learns to interpret F as \mathbf{f} .

The following definition contains the idea of the paper in a nutshell. Assume we are in stage $\text{Syn}_n = \langle E^n, S^n, H^n \rangle$. How do the new selection and interpretation strategies in the next stage Syn_{n+1} look like?

$$H^{n+1}(F) := \min\{\mathbf{f} \in \mathbf{Type} \mid \mathbf{f} \leq H^n(F) \wedge \|F\|_n \subseteq \llbracket \mathbf{f} \rrbracket_n\} \quad (4.3)$$

$$S^{n+1}(\mathbf{f}, \mathbf{f}') := \min\{F \in NL \mid \mathbf{f} \leq H^{n+1}(F) \leq \mathbf{f}'\}. \quad (4.4)$$

Where $\llbracket \mathbf{f} \rrbracket_n$ denotes the *extension* of \mathbf{f} in E^n , i.e. $\llbracket \mathbf{f} \rrbracket_n := \{e \in E^n \mid e : \mathbf{f}\}$; $\|F\|_n$ is the set of all entities where the speaker has in fact used F to classify them, i.e. $\|F\|_n := \{e \in E \mid \exists \mathbf{f}, \mathbf{f}' :$

$\langle e, \mathbf{f}, \mathbf{f}' \rangle \in E^n \wedge S^n(\mathbf{f}, \mathbf{f}') = F\}$. H^{n+1} and S^{n+1} describe both the hearer's and the speaker's learning. The hearer's learning precedes the speaker's, but we put both processes together in one stage. This learning should take place only with respect to actually used forms. If a form is never used, then the hearer can associate no restricted information with this form. Hence, we have to check which forms are used in each stage. We collect them in a set NL_{n+1} :

$$NL_{n+1} := \{F \in NL_n \mid \exists(e, \mathbf{f}, \mathbf{f}') \in E^n \ S^n(\mathbf{f}, \mathbf{f}') = F\} \quad (4.5)$$

If learning takes place with respect to NL_{n+1} only, then we have to restrict the definition of H^{n+1} in (4.3) to this set. The actual selection and interpretation functions H^{n+1} are defined by:

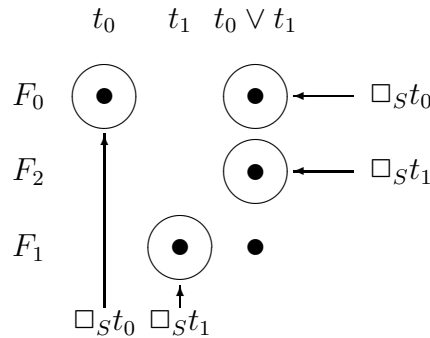
$$H^{n+1}(F) := \begin{cases} H^n(F) & \text{for } F \notin NL_{n+1} \\ H_*^{n+1}(F) & \text{else} \end{cases}, \quad (4.6)$$

where H_*^{n+1} is the function defined in (4.3). The diachronic model $(\text{Syn}_i)_{i=0,\dots,n}$ is totally determined by semantics and speaker's preferences on forms.

Let us apply this model to the *Kill-and-cause-to-die* Example (i)! The observed interpretations emerge as the result of a process involving two stages: (1) In the initial stage all killing events are direct killing events, i.e. in the first stage Syn_0 there are only events e which represent direct killings. The speaker will always use *kill* to denote these events. Hence, interpreters will learn that *kill* is always connected with direct killing. They *associate kill* with direct killing. The relevant types are $\mathbf{t}_0 = \text{direct killing}$ and $\mathbf{t}_1 = \text{indirect killing}$. Hence, we find $H^1(F_2) = \mathbf{t}_0$ and therefore the speaker will learn that hearers associate *kill* with direct killing. We observe further that the situation turns from a class III example into a class II example with $F_2 \prec F_3 \prec F_1$.

(2) In the second stage Syn_1 the speaker encounters an instance e' of an indirect killing. He has to avoid misleading associations and use the more complex form *cause to die*. We find that $S^1(\mathbf{t}_1, \mathbf{t}_0 \vee \mathbf{t}_1) = \min\{F \in NL \mid \mathbf{t}_1 \leq H^1(F) \leq \mathbf{t}_0 \vee \mathbf{t}_1\} = F_3$. He cannot select F_2 because $\mathbf{t}_1 \not\leq H^1(F_2)$. $F_3 \notin NL_1$, hence $H^1(F_3) = \mathbf{t}_0 \vee \mathbf{t}_1$. If we assume that the speaker always knows whether it was a direct or an indirect killing, then the hearer will learn that *cause to die* is always connected to an indirect killing \mathbf{t}_1 ; hence $H^2(F_3) = \mathbf{t}_1$. This in turn can be exploited by the speaker, and he will start to use *cause to die* for expressing \mathbf{t}_1 .

Let's turn to Example (v). We provide a graphical solution. The first row in the graph represents the speaker's possible intentions on how to classify an object. \Box_S is to be read as *the speaker knows that...* Hence, $\Box_S \mathbf{t}_0$ means that the speaker knows that the entity he classifies is of type \mathbf{t}_0 . The circles around bullets are to indicate that these form-meaning pairs are optimal according to his preferences. The arrows from $\Box_S \mathbf{t}_i$ indicate that this optimality depends on the speaker's knowledge. The situation for Case II examples looks as follows:



We can see that the speaker will use the general form F_2 only if he knows that the entity e has to be classified as being of type t_1 . Hence, as a matter of fact, if the hearer knows that the speaker knows the type of e , he can safely infer from an utterance of F_2 that the entity is of type t_1 . This explains why *Flüssigkeit* in (v) is interpreted as meaning *not water*.

So even a first survey shows how associative learning can lead to stronger interpretations and differentiations of meaning. Moreover, the survey provides us with a classification of utterance situations.

- [1] D. Beaver (2000): *The Optimization of Discourse*; ms. Stanford; to appear in *Linguistics and Philosophy*.
- [2] D. Beaver, H. Lee (2003): *Input–Output Mismatches in OT*; To appear in: R. Blutner, H. Zeevat (eds.): *Optimality Theory and Pragmatics*. Palgrave/Macmillan..
- [3] A. Benz (2001): *Towards a Framework for Bidirectional Optimality Theory in Dynamic Contexts*; ms., Humboldt Universität Berlin. Available as ROA 465-0901 and from <http://www.anton-benz.de>.
- [4] A. Benz (2003): *On Coordinating Interpretations - Optimality and Rational Interaction* ; To appear in P. Kühnlein, H. Rieser, H. Zeevat (eds.): *Perspectives on Dialogue in the New Millennium*; preliminary paper available from <http://www.anton-benz.de>.
- [5] R. Blutner (1998): *Lexical Pragmatics*; Journal of Semantics 15, pp. 115–162.
- [6] R. Blutner (2000): *Some Aspects of Optimality in Natural Language Interpretation*; In: Helen de Hoop & Henriette de Swart (eds.) *Papers on Optimality Theoretic Semantics*. Utrecht Institute of Linguistics OTS, December 1999, pp 1-21. Also: *Journal of Semantics* 17, pp. 189-216.
- [7] R. Blutner, G. Jäger (2000): *Against Lexical Decomposition in Syntax*; In A.Z. Wyner (ed.): *Proceedings of the Fifteenth Annual Conference, IATL 7*, University of Haifa, pp. 113-137 *Proceedings of IATL 15*, University of Haifa.
- [8] B. Carpenter (1992): *The Logic of Typed Feature Structures*; Cambridge University Press, Cambridge.
- [9] P. Dekker, R. v. Rooy (2000): *Bi–Directional Optimality Theory: An Application of Game Theory*; *Journal of Semantics* 17, pp. 217–242.
- [10] L. Horn (1984): *Towards a new taxonomy of pragmatic inference: Q–based and R–based implicature*; In: D. Schiffrin (ed.): *Meaning, Form, and Use in Context: Linguistic Applications*, Georgetown University Press, Washington, pp. 11–42.
- [11] G. Jäger (September 2000): *Some Notes on the Formal Properties of Bidirectional Optimality Theory*; ms, ZAS Berlin.
- [12] D. Lewis (1969): *Convention*; Harvard University Press, Cambridge.
- [13] J. Mattausch (November 2000): *On Optimization in Discourse Generation*; master thesis, Universiteit van Amsterdam.
- [14] P. Parikh (1990): *Situations, Games and Ambiguity*; In R. Cooper, K. Kukai, J. Perry: *Situation Theory and its Applications I*, CSLI Publications, Stanford.
- [15] P. Parikh (2001): *The Use of Language*; CSLI Publications, Stanford.
- [16] R. van Rooy (2002): *Signalling Games select Horn Strategies*; ms Universiteit van Amsterdam; to appear in *Linguistics & Philosophy*.

CONSTRAINTS IN OT: a comparison between unidirectional and bidirectional OT

Hanneke van der Grinten

University of Nijmegen

Abstract. Bidirectional OT integrates the production perspective of OT Syntax and the interpretation perspective of OT Semantics. In this paper I will investigate the bidirectional OT of Blutner (2000), who formalizes conversational principles in order to explain blocking phenomena as the *Division of Pragmatic Labor*. I will show that the incorporation of conversation principles in bidirectional OT is in many respects problematic. I will argue that Blutner's theory deviates from its unidirectional counterparts in some essential features, and that its explanatory force is limited to a rather small domain.

1. Unidirectional OT

It is generally known that in OT two perspectives can be taken: the perspective of the speaker, who selects the optimal form given a certain meaning, and the perspective of the hearer, who selects an optimal interpretation given a certain form. The first perspective is reflected in OT Syntax, the second in OT Semantics. Both OT accounts I will refer to as *unidirectional OT*, in contrast to *bidirectional OT*, which will be introduced later. The initial account of OT Semantics (Hendriks & De Hoop, 2001) makes predictions about preferred interpretations on the basis of a set of ranked constraints. The basic idea is the following. Each utterance can be seen as having a (possibly infinite) set of possible interpretations. This set is submitted to a set of ranked constraints, whose ranking is empirically determined. Most interpretations will violate one or more constraints. The number of violations of higher ranked constraints then determines which of the *possible* interpretations is evaluated as the *optimal* interpretation of that utterance. The result of this evaluation is immediately visible in the standard tableau notation: the optimal interpretation is the one that does not violate the highest ranked constraint violated by competing candidates.

The procedure in unidirectional OT consists of two steps: (1) determination of constraint violation and (2) evaluation of possible outputs. This can be illustrated with the following example:

- (1) Often when I talk to a doctor, the doctor disagrees with him. (Hendriks & De Hoop, 2001)

In interpreting this utterance, two constraints are supposed to be at work:

1. DOAP: Don't Overlook Anaphoric Possibilities. Opportunities to anaphorize text must be seized.
2. Principle B: If two arguments of the same semantic relation are not marked as being identical, interpret them as being distinct.

The ranking of these constraints is Principle B >> DOAP.

The first step is submission of the possible outputs (interpretations) to these constraints, which gives the following result:

Input	output	Principle B	DOAP
Often when I talk to <i>a doctor</i> , <i>the doctor</i> disagrees with <i>him</i>	a doctor ₁ the doctor ₁ him ₁	*	
	a doctor ₁ the doctor ₂ him ₁		*
	a doctor ₁ the doctor ₁ him ₂		*
	a doctor ₁ the doctor ₂ him ₂	*	*
	a doctor ₁ the doctor ₂ him ₃		**

The second step is evaluation of the tableau, where ! shows fatal constraint violations, and ⇒ selects the optimal output(s), as shown in tableau 2:

Input	output	Principle B	DOAP
Often when I talk to <i>a doctor</i> , <i>the doctor</i> disagrees with <i>him</i>	a doctor ₁ the doctor ₁ him ₁	*!	
	⇒ a doctor ₁ the doctor ₂ him ₁		*
	⇒ a doctor ₁ the doctor ₁ him ₂		*
	a doctor ₁ the doctor ₂ him ₂	*!	*
	a doctor ₁ the doctor ₂ him ₃		**!

Two features of this procedure must be stressed:

- (a) the outputs are submitted to the constraints *in isolation*, i.e. without taking alternatives into account.
- (b) it is only in the selection of optimal candidates that the alternatives are compared to one another; the selection procedure is clearly of another, higher level than the determination of constraint violation.

On these points unidirectional OT deviates from its bidirectional counterpart.

2. Bidirectional OT

Although unidirectional OT Semantics accounts for preferences of interpretation, it cannot account for so called *blocking phenomena* (Blutner, 2000). Blocking examples show that not only the production of an utterance (its form) affects the interpretation, but also the other way round: certain forms are *blocked* because the intended interpretation can be described more economically by using an alternative expression. An example of complete blocking is (2):

(2) ?The table is made of tree.

The use of *tree* is blocked because of the existence of the more specific alternative *wood*, as in (3):

(3) The table is made of wood.

Although this is a case of *complete blocking*, as an utterance with the non-economical expression (in this case *tree*) usually gets no interpretation at all, it seems to be of a pragmatic nature, as there are contexts in which (2) might be said felicitously, as in (4):

(4) A: This table doesn't contain any living material
B: (Of course it does!) The table is made of tree.

Apart from *complete* blocking as in (2), there are also cases of *partial* blocking, in which the non-economical expression is only blocked for a certain interpretation, i.e. the interpretation which refers to the most stereotypical situation. In this case the non-economical form gets *another* meaning:

(5) I caused the poor rabbit to die.

The speaker could have said (6) instead of (5):

(6) I killed the poor rabbit.

An utterance of (5) is clearly non-economical (or *marked*). Where (6) will lead to the interpretation of direct killing, this interpretation is excluded for (5). As a result (5) will be interpreted as an act of indirect killing. Partial blocking leads to the effect that *unmarked forms tend to be used for unmarked situations and marked forms for marked situations* (Horn, 1984: 26). This effect is known as the Division of Pragmatic Labor (Horn, 1984).

In order to account for the Division of Pragmatic Labor Blutner (Blutner, 2000) develops a *weak* bidirectional Optimality Theory. It is this weak version to which I will pay attention here. In this bidirectional framework, Blutner makes use of pragmatic principles which are widely held to govern conversation. These principles originate with Grice and have been reformulated by Horn (Horn, 1984) and Levinson (Levinson, 2000). Horn and Levinson both reduce the number of principles.

There are a couple of problems connected to Horn's and Levinson's use of these principles. I will give a short summary, in order to make clear what I think is the best concept of these principles.

2.1 Three conversational principles: *Q*, *I*, and *M*

Q is a *Quantity* principle, and is responsible for implicatures based on the informativeness (scalars/clausals). It forces the speaker to be as informative as possible. For example:

- (7) I corrected some of the mistakes in my paper.
- (8) I corrected all of the mistakes in my paper.

As ‘all’ is a stronger (thus more informative) expression than ‘some’, the hearer can infer from (7) that I did not correct all of the mistakes, for otherwise I would have said (8) instead, in order to satisfy *Q*. The hearer perspective of the *Q* principle is thus to infer that the stronger expression does not hold.

Horn manipulates this informativeness principle to account for implicatures based on the form of an utterance. This, as we will see, is also the case in Blutner’s treatment of the conversational principles.

I (called *R* by Horn) is a *minimization* principle. Although it originates from one of Grice’s Quantity principles, it is used, both by Horn and Levinson, to instruct speakers to minimize the *informative content* of an utterance as well as the *form* of the utterance. The hearer perspective of the *I* principle is *enrichment*: enrich the speaker’s utterance up to the most coherent and stereotypical interpretation. For example:

- (6) I killed the poor rabbit.

As (6) is a minimal expression it must be enriched with stereotypical information: I killed the rabbit directly, with my own hands.

M is a *form* (or *Manner*) principle, which we don’t find in Horn’s theory. This principle instructs a speaker to use a marked form in order to refer to a marked or non-stereotypical situation. The hearer perspective of the *M* principle is to interpret a marked form as referring to a marked or non-stereotypical situation. For example:

- (5) I caused the poor rabbit to die.

As (5) is a marked expression it must be interpreted as referring to a non-stereotypical situation: I killed the rabbit in an indirect way, eg. by not giving him any food.

From this it follows that the Division of Pragmatic Labor is *stipulated* by the *M* principle. Another stipulation in Levinson’s theory is the hierarchy of the *Q*, *I* and *M* principles which guarantees that if two principles are in conflict with each other, the highest ordered principle will “win”, i.e. will bring about the “potential” implicature. Levinson’s hierarchy is as follows:

$Q > I$
 $Q > M$
 $M > I$

The ranking of the *Q*/*I* and *Q*/*M* principle will not be taken into consideration, as they play no role in bidirectional OT. The hierarchy of *M*/*I* is of a different nature compared to the other two ordering relations. This time it is not the case that there are two “potential” implicatures, (a potential *I*-implicature and a potential *M*-implicature) which

are in conflict, nor that by the ordering relation one can tell which of the potential implicatures will be brought about. By contrast, M is supposed to be working as a *blocking* mechanism: it blocks the interpretation to the stereotype, and thus brings about the opposite, i.e. an inference to the non-stereotypical situation. This only has its influence on the hearer: if a speaker violates I (and thus uses a marked expression) any I-implicature (that is, any inference to the stereotype) is blocked by M. The M principle will bring about that the hearer infers to the non-stereotypical situation. This is the way in which the hierarchical ordering of I and M must be understood: we assume that speakers violate I in order to satisfy M.

Blutner claims that bidirectional OT accounts for the Division of Pragmatic Labor without stipulating it, as the M principle does, and without any stipulation of ordered principles, because the hierarchy follows automatically from the theory.

In the next section I will analyze his theory in order to show that:

- (a) Blutner does not convincingly intergrate the conversational principles in OT.
- (b) Constraints in bidirectional OT have a *relative* character. As a result the first step of the OT-procedure already is an evaluative one. The procedure in bidirectional OT thus deviates from the procedure in unidirectional OT.
- (c) In its present form, Blutner's bidirectionality is limited to markedness phenomena although Blutner claims that the explanatory force of his theory is extended to other phenomena.

3. Conceptual analysis of Blutner's theory

Bidirectional OT is formulated in terms of the above-mentioned principles Q and I. In a more transparant formulation (Jäger, 2001) this definition can be formulated as follows¹:

A form-interpretation pair, in which A and A' are coextensive forms, t and t' are interpretations, is optimal iff:

Q: there is no other optimal pair (A',t) such that $\langle A',t \rangle > \langle A,t \rangle$

I: there is no other optimal pair $\langle A,t' \rangle$ such that $\langle A,t' \rangle > \langle A,t \rangle$

where $>$ means 'more harmonic/economical'.

Informally, this means roughly that **Q** selects the most economical form for expressing a given interpretation, **I** selects the most coherent interpretation for a given form. At first sight the procedure in bidirectional OT seems to be equivalent to the procedure in unidirectional OT, except for the fact that this time the candidates to be evaluated are *form-interpretation pairs*, instead of *interpretations*. The first step in bidirectional OT is to determine violations of constraints, which are represented, as usual, by decorating an OT tableau with asterisks. The second step is to evaluate the alternatives and to determine optimal form-interpretation pairs. In bidirectional OT, however, this second step is governed by **Q** and **I** so that we have to distinguish between two sorts of constraints:

- a) the constraints *in* the OT tableaux, which are the 'normal' constraints we also find in the unidirectional account.

¹ I will use the font **Q** and **I** to distinguish Blutner's use of these principles from the original principles themselves.

- b) the constraints with which the tableaux are evaluated, i.e. **Q** and **I**, which I will call *meta-constraints*. These are not found in the unidirectional framework.

In order to deal with the Division of Pragmatic Labor Blutner formulates two ‘normal’ constraints: one constraint F is a constraint on linguistic forms which ‘collects the effects of linguistic markedness’, while C is a constraint on resulting contexts which ‘refers to coherence and informativeness’.

Determining constraint violations of example (5) and (6) gives the following result in a bidirectional tableau. This is the first step, as in unidirectional OT:

Form ↓		F	C		F	C
<i>killed</i>						*
<i>caused to die</i>		*			*	*
Interpretation →		‘direct killing’			‘indirect killing’	

Second evaluation of the tableau, which is governed by **Q** and **I**, shows the optimal candidates:

Form ↓		F	C		F	C
<i>killed</i>	⇒ ☞					*
<i>caused to die</i>		*		⇒ ☞	*	*
Interpretation →		‘direct killing’			‘indirect killing’	

In spite of the superficial resemblance with unidirectional OT, there are some essential differences between the two accounts.

The constraints in unidirectional OT are pragmatic constraints formulated as concrete maxims (cf. ‘Don’t Overlook Anaphoric Possibilities’, or: ‘If two arguments of the same semantic relation are not marked as being identical, interpret them as being distinct’). These constraints are not *ad hoc* invented in order to get the desired results, by contrast they are externally motivated constraints. The constraints F and C in Blutner’s OT tableau lack concrete instructions, which makes it hard to tell *what exactly* they measure. There is a strong suggestion that these constraints divide the possible forms into *marked* versus *unmarked* and the possible interpretations into *coherent/informative* versus *incoherent/non-informative*. That would mean that F is violated if a speaker uses a marked form instead of an unmarked, and that C is violated if a hearer selects a non-stereotypical interpretation instead of a stereotypical one. In the search for an external motivation for these constraints, we can hardly think of any motivation other than the conversation principles as formulated by Horn and Levinson. It thus seems to be appropriate to say that:

F = choose unmarked form, which is the same as speaker principle I (or Horn’s R)

C = choose unmarked interpretation, which is the same as hearer principle I, not explicitly formulated by Horn.

As the ‘normal’ constraints seem to be based on the I principle, it seems unlikely that the meta-constraints are based on it as well. Moreover, the Q principle doesn’t seem to play a role whatsoever, as it is an informativeness principle and as such cannot select the most *economical form*.

The correspondance between F and C on the one hand and I on the other, also lays bare the mutual dependency of form-interpretation pairs, for there is no absolute standard for a form to be called ‘marked’ or an interpretation to be called ‘coherent & informative’. Markedness, coherence and informativeness are *relative* notions. A form is *marked with respect to another form* or *marked to a certain degree* and it is fairly clear that no clean borderline can be drawn between marked/unmarked. The same holds for interpretation: an interpretation is more or less coherent/informative than another. The working of the conversation principles is based on the *choice* a speaker has between various forms to express various situations. The knowledge of available expression-alternatives is indispensable for the working of these principles. In OT this means that in these cases constraint violation can only be determined by taking the alternatives into account: a certain form violates F *with respect to another form*. This thus deviates from unidirectional OT in which the possible interpretations are all *in isolation* submitted to the set of ranked constraints. Whether or not a certain interpretation output violates a certain constraint is independent of the available alternatives. Only by evaluating the possible interpretation are the outputs compared to one another. In bidirectional OT the *first* step already has an evaluative character, as the possible candidates are not submitted to constraints *in isolation* but *are compared with one another*.

I have shown that the procedure in uni-and bidirectional OT is of a different kind. In the following I will make plausible that what Blutner calls Q and I, takes over the working of the M principle. First it must be noted, however, that of course we could add the M principle to the set of constraints. This, however, would mean that there is no point in working this out in bidirectional OT: OT would not give any formal reduction to the problem and the Division of Pragmatic Labor would then being stipulated. The stipulated ranking between I and M would also remain intact. The reason why Blutner formulates his meta-constraints is precisely to formalize and reduce the problem. Although this formalization seems in fact a reduction of what Horn and Levinson do, the working of these meta-constraints, on which his theory is founded, turns out to be rather limited. It plays no role in giving explanations for phenomena other than markedness, although Blutner presents his account as a *general* theory for which a definition in terms of **Q** and **I** is indispensable. This brings me to the more fundamental question what exactly *is* bidirectionality, and whether conversational principles should be done by OT. To these questions I will turn now.

4. Bidirectionality

We saw that Blutner defines his bidirectional theory in terms of **Q** and **I**, which I have called *meta-constraints*. These constraints work as an evaluation mechanism, to select optimal form-interpretation pairs. A similar evaluation system is not present in unidirectional OT Semantics. The question is whether it is a natural feature of bidirectionality to have such an evaluative system. I think it is not. The essence of bidirectionality is to account for the fact that interpretational preferences can affect a speaker’s utterance. Ordinarily speaking, bidirectionality shows that speakers search not only the best expression *regarding their own perspective*, they also take into account *what is better from a hearer’s perspective*. In OT this means that the speaker’s choice of expression, can be influenced by constraints which are not only constraints on *forms*, but constraints on *form-interpretation pairs*. It is thus that certain expressions can be blocked, because it is better *from a hearer’s perspective* that the intended interpretation

is conveyed by means of an alternative expression. That is the essence of bidirectionality. Contrary to what Blutner claims, a definition in terms of **Q** and **I** is superfluous in these cases. Consider for example (Blutner, 2000:211):

- (9) A: Did you hear about John?
 B: No, what?
 (a) A: He had an accident. A car hit him.
 (b) A: He had an accident. ??The car hit him.

Blutner's explanation of the infelicitousness of (9b) is that it is blocked by the more economical utterance (9a), due to the fact that *the car* must be accommodated, while *a car* need not. Because of this blocking it cannot be interpreted properly. But contrary to what Blutner says, no meta-constraints are necessary to explain these facts in OT, which can be shown by drawing a tableau. For the sake of simplicity I will consider just one constraint. The crucial point in this example, is that this constraint, *Avoid Accommodation*, is a *hearer based* constraint, and yet affects the speaker's choice to utter *a car* instead of *the car*.

↓ Form	avoid accommodation
<i>a car</i>	
<i>the car</i>	*
interpretation →	'a newly introduced car'

Evaluation selects the optimal candidate, without falling back on any evaluative system as it can be read immediately from the tableau:

↓ Form	avoid accommodation
<i>a car</i> ⇨	
<i>the car</i>	*!
interpretation →	'a newly introduced car'

We can now turn to the last point, i.e. showing that Blutner's definition of bidirectionality in terms of **Q** and **I**, is limited to explain the Division of Pragmatic Labor.

4.1 **Q** and **I**

Apart from partial blocking, as is the Division of Pragmatic Labor, I am not aware of any case for which a bidirectional OT with a selection mechanism in terms of **Q** and **I** is indispensable. That makes me wonder what exactly the status of these meta-constraints is. My suggestion is that Blutner's **Q** and **I** correspond in fact for a large part with the both sides of the M principle, although in a better and more elegant formulation. Instead of the speaker's instruction "choose a marked form to express a marked situation", **Q** (which I take to be its counterpart in bidirectional OT) selects the most economical form *which is not optimal with another interpretation*, in other words: if the most economical form is optimal to express a simpler interpretation than you want to express, this form

can *not* be optimal for the (*marked*) interpretation you want to express, and as a consequence the speaker has to use the less economical expression. The correspondance with Levinson's M principle is clear, although in Blutner's terms it is possible to avoid some problems Levinson has to deal with. It thus seems that the stipulation of a hierarchical ordering of M and I is not only present in Levinson's theory, but in Blutner's bidirectional OT as well, as he places **Q** and **I** (which I see as speaker- and hearer-perspective of the M principle) on a *higher level*. The result is that **Q** and **I** only have a function in selecting an optimal candidate in case F and C (= the I principle) are violated: that is exactly the same as we explained in the working of I versus M: M is higher in the hierarchical ordering (in OT: M is on a meta-level) because speakers *violate* I in order to *satisfy* M.

5. Conclusion

I have tried to show that the meta-constraints in terms of which Blutner's weak version of bidirectional OT is formulated are no essential feature of bidirectional OT. By contrast, these meta-constraints seem to take over the task of the M principle, in case the form-interpretation pairs are submitted to constraints like F and C, which origins probably lay in the I principle. Although Blutner shows an elegant way to deal with the Division of Pragmatic Labor, it is just a minor reduction of the original conversational principles: the meta-constraints only act when the constraints F and C (= the I principle) are violated. Violation of both F and C means satisfying M, and thus satisfying **Q** and **I**. Levinson's hierarchical ordering of I and M is thus present in bidirectional OT by using constraints on two different levels. Bidirectionality in terms of **Q** and **I** is relevant only in a rather limited domain, for in most cases we get the right results without making an appeal to these meta-constraints. Apart from this, it is shown by others that Blutner's weak bidirectional OT falls short of explaining other blocking phenomena than partial blocking, and overgenerates in a lot of cases (Beaver & Lee, 2003). The question remains whether OT is the appropriate way to deal with conversational principles at all.

References

- Beaver, D. and Lee, H., 2003. Form-Meaning Asymmetries and Bidirectional Optimization. In: R. Blutner and H. Zeevat (eds.), *Pragmatics in OT*. Palgrave MacMillan
- Blutner, R., 2000. Some aspects of Optimality in Natural Language Interpretation. In: *Journal of Semantics* 17: 189-216. Oxford University Press.
- Hendriks, P. and Hoop, H. de, 2001. Optimality Theoretic Semantics. In: *Linguistics and Philosophy* 24: 1-32.
- Horn, L.R., 1984. Toward a New Taxonomy for Pragmatic Inference: Q-based and R-based Implicature. In D. Shiffrin (ed.), *Meaning, Form, and Use in Context* (pp. 11-42). Washington DC: Georgetown University Press.
- Jäger, G., 1999. Optimal Syntax and Optimal Semantics. Handout for talk at DIP colloquium 1999.
- Levinson, S.C., 2000. *Presumptive Meanings. The Theory of Generalized Conversational Implicature*. Cambridge, MA: The MIT Press.

Markedness and Economy on Signs

Henk Zeevat

University of Amsterdam

1. Introduction

This paper introduces a notion of economy on linguistic signs that comes in place of notions like optimal form, optimal meaning, optimal form meaning pair and the like. These notions can all be defined in terms of maximally economical signs. An optimal form (for an input) is a form associated with the input in a maximally economical sign, an optimal meaning (for a form) is a meaning associated with the form in a maximally economical sign and an optimal form-meaning pair is the pair consisting of the form and the meaning in a sign.

The current way of looking at optimality theoretic syntax, semantics and pragmatics also addresses another goal, the relation between optimality theory and sign based semantics such as practiced in frameworks like HPSG and—in a perhaps unconventional understanding of that enterprise— of LFG.

And it contributes to another goal as well, since it gives a theory underlying the use of statistical methods in natural language processing. Statistics is the key to understanding the different economy dimensions because we can equate the most likely components in a sign—given the rest of the sign—with the most economical elements.

I will focus in this paper on explaining the notion and on the relation with optimality theory and will try to answer three questions. Why is this an improvement of bidirectional optimality theory? Can it still be interpreted as a kind of optimality theory? Can it still form the basis for functional-historical accounts of aspects of language along the lines of Zeevat & Jäger (2002) and Jäger (to appear)?

A sign is here the combination of a linguistic form with a linguistic meaning and an association: a relation between the components of the linguistic form and the components of the linguistic meaning.

Economy falls apart into three different notions, one relating to form, the other to meaning, one related to the association between them. But all three notions essentially involve the other two components: the form is most economical as a form associated in this particular way to the meaning. And the meaning is most economical as a meaning associated in this way to the form. And the association is most economical as an association between two maximally economical components. We will run through the different aspects of economy, but assuming that this will make sense, the general theory is simple.

- (1) A sign is economical iff there is not another correct sign that associates a more economical form to the same meaning or associates a more economical meaning to the same form or another association between its form and meaning that is more economical

2. Problems in Bidirectional Optimality Theory

Iconicity is the principle that complex meanings get long expressions and simple meanings simple ones. There are two quite different phenomena that fall under this principle. The first is a historical and statistical phenomenon. Simple and frequent meanings tend to be expressed by short expressions, whereas complex and rare meanings tend to be expressed by longer ones. Zipf's law (Zipf (1949)) describes the relationship and, presumably, there is some fact about the evolution of language use that is responsible for the phenomenon[‡]

The other phenomenon is the effect of Grice's maxim *Be brief* a phenomenon that is exemplified by Horn's famous opposition:

- (2) a. Black Bart killed the sheriff
b. Black Bart cause the sheriff to die.

and the no less famous example from Grice (1975):

- (3) a. Mrs. T produced a series of sounds closely resembling the score of "Home Sweet Home".
b. Mrs. T sang "Home Sweet Home".

The explanation of the two phenomena cannot be the same. There may be a historical process that associates simple meanings with short forms but it cannot apply in this case where the marked forms have an extremely low frequency and therefore cannot possibly have acquired their marked meaning by associating to it through an evolutionary process. There must be another way in which the complex form acquires the marked meaning. Taken literally "cause to die" is just "kill" and singing a song is the production of sounds that closely resemble the score of that song. The emergence of a marked meaning seems due both to *Be brief* and the existence of the possibility to be brief, i.e. the shorter form. Notice also, the difference between these cases and historical iconicity. In the historical case, we have a conventional association between the long form and its meaning and the meaning can be spelt out as well as the meaning of other words. In pragmatic iconicity, the complex meaning is vague: there is something unusual about the killing and the singing. Indirect ways in which Black Bart killed the sheriff would do, but indirectness is not part of the assertion, we may interpret the speaker as saying that Mrs. T did not do a very good job in her rendering of "Home Sweet Home", but negative evaluation is not part of the conventional meaning. This is the sort of vagueness that for Grice is the hallmark of conversational implicature. Notice also that any particular effect of the marked form can be cancelled. If indirectness would be conventionally associated with the marked form in the Horn example or esthetical disapproval in the Grice example, the examples in (4) would be inconsistent and not just a little bit enigmatic (we need a reason for the marked form and some reasons are ruled out now).

- (4) a. Black Bart caused the sheriff to die in a very direct way.
b. Mrs. T produced a series of sounds closely resembling the score of "Home Sweet Home" and did so beautifully.

In Blutner (2000), Reinhard Blutner introduces a general approach to both optimality theory and the abstract pragmatics of Horn(1984) and Levinson (2000) in terms of the Q-, I- and M-principle. His solution has a relation to Optimality Theory if one takes optimality theory as a general theory of markedness. OT-competitions among forms for a particular meaning using a constraint system *S* order the different forms on a markedness dimension (a preordering). OT-competitions among meanings in an interpretational OT using the same

[‡] Frequency of a word makes its recognition easier, and thereby the functional pressure to realise all its phonological features correctly. This leads to unchecked variation and presumably to loss of the features that are only optionally realised. Loss of features leads to shorter words.

constraint system S can be seen as defining a markedness relation among meanings. In this interpretation, OT syntacticians, semanticists and pragmatists are all concerned with isolating the principles that determine what forms are marked in a particular language or with principles that make interpretations marked. It is a collaborative effort since OT-syntax normally bases itself on some concept of the input to the syntactic competition and theories of the input naturally are a characterisation, at some level of abstraction, of what the speaker wants to say. Semantics and pragmatics are concerned with exactly the question of what the speaker wanted to say with her utterance, but then want to relate that to explicit characterisations of speech acts and contents. The difference between concepts of the input and semantics/pragmatics is mainly due to the different demands on the representation: something that can be manipulated by theorem provers versus something that is rich enough to serve as a basis for comparing syntactic forms. Smolensky (1996) and Tesar & Smolensky (2000) give arguments for assuming the same system of constraints in production and interpretation.

Armed with the two preorders $<_{syn}$ and $<_{sem}$, derived from the constraint system S , it is possible to define optimal forms for a meaning and optimal meanings for a form in one single definition of superoptimal pairs.

A pair $\langle m, f \rangle$ is superoptimal iff there is no superoptimal pair $\langle m', f \rangle$ or $\langle m, f' \rangle$ where $m' <_{sem} m$ or $f' <_{syn} f$.

Using OT systems with finite constraints and a partial order, $<_{syn}$ and $<_{sem}$ are both well-founded preorders and therefore the definition of optimal pair is a proper recursive definition, as shown in Jäger (2002).

With some charity, optimal pairs gives the solution to the problem in the previous section. Assume that meanings can come into two flavours, the vanilla meaning and the cherry meaning. Cherry indicates that there is more to be told, vanilla is the default case. We further assume that forms are compared by **Economy**, a principle that compares the length of forms. The pair $\langle kill, vanilla \rangle$ is optimal because there is no shorter form or less marked meaning. $\langle kill, cherry \rangle$ is not optimal because there is a less marked meaning in the optimal pair $\langle kill, vanilla \rangle$. $\langle kill, vanilla \rangle$ also successfully eliminates $\langle cause\ to\ die, vanilla \rangle$: *cause to die* is too long. So what remains is $\langle cause\ to\ die, cherry \rangle$.

I am not happy with this way of deriving the effect. Both of the assumptions needed are rather suspect: cherry and vanilla meanings and **Economy**. It seems that the cherry meanings arise as part of the effect and cannot be presupposed. It is also questionable that a simple comparison of the length of expressions is sufficient. Using frequency instead of length gives the same results (the more frequent expression alternative is the unmarked one). There are also many cases (e.g. drink vs. have a drink, stop the car vs. make the car stop) where no effect can be observed from extra length.

There are also other things wrong with Blutner's concept of optimal pairs. Some of the problems are inherent to any bidirectional system, like versions of the Rad-Rat problem (Hayes & Hayes (1989)) emerging in syntax. Here OT principles like **Faith** or **Stay** can be responsible for systematically eliminating meanings (the Rad-meaning for the pronunciation /rat/, the object interpretation for *Welches Mädchen liebt Peter?*) that —as intuition tells us— are just there.

Another problem is pointed out by Lee & Beaver (to appear), the problem that any suboptimal candidate becomes optimal after a number of recursive rounds. The problem seems sufficient for giving up on superoptimality.

One is tempted to say that perhaps this form of bidirectional OT is only suitable for historical explanations (e.g. Blutner p.c.). But if we accept that conclusion, we also have a substantial problem. We had an account of pragmatic iconicity and it cannot be recovered in the historical account that replaces it. The bidirectional account seems to recover Grice's plausible explanation of what is going on in these examples.

3. Economy

Before we used OT constraints to define $<_{syn}$ and $<_{sem}$. The plan is now to make these into primitives of the theory and see what happens. Only later, we will return to the constraints.

I will first try to define semantic economy § in terms of the least change that the speaker proposes to make to the common ground as it stands.

The least marked case is the case of no change at all, not even a change to the current focus of attention. Attentional changes are slightly more marked but still give no new information. The information supplied has all been supplied before and integrated in the common ground. Information change that just consists in adding new information is the normal case, where a distinction can be made between adding new objects and new information about objects. In the file card metaphor (Karttunen (1976)), this is the distinction between writing new information on the cards and adding new cards to write information on. Another distinction on this level, is between adding new objects that are functionally related to old objects and adding new objects without any relation to old objects. The most marked case is adding new information as a replacement of old information, as happens in corrections. (5) recapitulates of these observations.

- (5) old information in focus of attention <
old information <
new information about old <
new related object <
new unrelated object <
correction <
correction on object

Typically this hierarchy can be applied locally under the association with the surface form. It then expresses general preferences and gives for example the natural preferences for interpreting definite descriptions (or in other languages lacking definiteness markers for bare NPs).

But also for e.g. personal pronouns and indefinite NPs. We need a way in which the current model can make indefinite NPs interpreted as referring to old objects more marked than interpretations in which they refer to new objects or to make new interpretations of personal pronouns marked. Taking NPs here as the example is an accident: the same points can be made using temporal objects, which are also preferably linked or resolved. I take this to be a question of probabilities and conventions. The marked interpretations are improbable given the language use and thereby marked. Probability is a separate source of markedness and allows us to have word meanings and meanings of morphemes and constructions. I refer to probability here rather than convention because it is the more basic case. Convention emerges from probability. Markedness can be seen as low probability of the occurrence. Avoiding marked meanings and marked forms strengthens the adherence to a convention if the probabilities favour it.

It is not clear to me that on the level of the surface form we also have a natural notion of markedness or that everything must be relegated to probabilistic patterns. Candidates for natural markedness would be only two principles. One is phonological complexity or length of expression, the other the coherence of constituents. But it can be maintained that also

§ In Zeevat (2001) I presented a constraint system for pragmatics. That system falls out of the markedness hierarchy given here, with corrections violating **Consistency** and new information violating ***Accommodation**. **Relevance** should come out as a preference for linked information, in this case linked to the “questions under discussion” in the context.

these arise from probabilities, as is the case with facts about word order and morphological marking.

The natural economy dimension is relational economy on pairs. We assume that we have a form-meaning pair with an association. Associations are arbitrary but would ideally be such that each semantic object is associated with overt elements in the surface form and each word with an element of the semantic representation. We must allow for probabilistic and conventional factors also here.

- (6) An economical sign is one that is not blocked by an economical sign that is less marked in one of the three dimensions.

The Grice/Horn examples can now be explained in the following way. If *Black Bart caused the sheriff to die* would just mean that Black Bart killed the sheriff, it would not be the surface form of a proper sign, since the verbal group can be replaced by the less marked *killed*. It therefore does not mean just that and leads to the pragmatic implicature that there was something special about the killing. This extra special feature can be indirectness, but it can also be something else (e.g. Black Bart did not know that the sheriff was hiding behind the sack of wool he was using for target practice.) The Grice example has the same explanation. The implicature is part of the recovery operation: it allows us to consider an uneconomical sign as economical.

Blutner also treats the semantics of “older gentleman” and of “not unhappy” in bidirectional OT. Older gentlemen are not young but are not very old either, even though the semantics of “older” seems to allow that. After all, if you are a properly old gentleman, you definitely belong to the group of older ones among the gentlemen. The explanation (both in Blutner and here) is simple blocking. Old would be the less marked form if the gentleman would be just old. We can apply the same reasoning to the case of “not unhappy”. Assume that people can be divided by their degree of happiness in the following five classes: properly unhappy, a bit unhappy, neither happy nor unhappy, a bit happy and properly happy. Semantically, “not unhappy” rules out the first two classes and leaves open the other three. Blocking then rules out the last class: properly happy, because a less marked expression is available there. The normal not unhappy person is then more happy than unhappy. It is not necessary to work with 5 semantic possibilities here, since the example can run directly with probability. Happy assigns probabilities to degrees of happiness with lower probabilities for extremely high happiness and lower positive degrees of happiness. This makes “not unhappy” assign higher probabilities to the areas where the degree is not negative but not high.

The Rat/Rad problem is a problem in OT phonology but can be reconstructed here. An OT production competition always produces the pronunciation /rat/ for the two Dutch words, because of word final devoicing overriding faithfulness with respect to voice in German or Dutch. The wrong prediction is that in an OT comprehension competition, the meaning *Rat* always wins from the meaning *Rad* because the latter and not the first transgresses the faithfulness constraint with respect to voice.

We do not have this problem here, since the meaning *Rad* is not more marked than the meaning *Rat* under the association with the pronunciation /rat/, (except possibly by the frequency of its occurrence). The pronunciation /rad/ for *Rad* is more marked because in German/Dutch pronunciation voice never occurs in word final position. Faithfulness constraints do not seem to play a role at all in the system that I have sketched so far and this is maybe a problem. How can we disallow the pronunciation of *Rat* as e.g. /rot/ without them?

It would seem that this is just the same question as lexical meaning. There is an abstract entity for the language user that is linked to pronunciation and perceptual properties in an

essentially conventional way. The abstract d is associated with two conventions: $|d\#|$ going to $/t\#/$ and d going to $/d/$ with the first overriding the second. In other cases a single convention will do. $|d|$ is not a natural feature anymore (the source of a phoneme) because of the two rules that realise it differently.

Lee & Beaver (to appear) notice that unrestricted weak bidirectionality leads to the absurd consequence that any form however bad will receive a meaning after all the forms that are not as bad will have been given meanings by bidirectional optimality theory provided that there are enough meanings around. Having enough meanings is a not uncommon situation in descriptive work using OT constraints, we normally assume a competition between all possible meanings. The example they provide of Korean case assignment and word order is representative. The point is that it does not happen: the prediction does not match the intuition about what goes on in Korean. In particular, there are a number of forms that are just not grammatical, even though they are predicted to express ever more marked meanings.

Our economy notion is inspired by weak bidirectionality and might therefore suffer from the same problem. The following constraint system is assumed: $*\text{subj}/\text{acc} > \text{Head-R} > \text{SO} > *\text{subj}^{\text{new}} > *\text{obj}^{\text{given}}$. $*\text{subj}/\text{acc}$ makes it bad that subjects are not nominatives, Head-R wants to have the verb last, SO wants to have the subject before the object, and the last two make non-topic subjects and non-focus objects bad. The first three constraints are markedness criteria on the syntax, the last two on the semantics. This gives us a linear markedness ordering on syntactic forms:

$$s_{\text{nom}} \text{ O } v < \text{O } s_{\text{nom}} v < s_{\text{nom}} v \text{ O } < v s_{\text{nom}} \text{ O } < v \text{ O } s_{\text{nom}} \text{ O } < \text{S O } v < \text{O S } v < \text{S v } < v \text{ S O } < v \text{ S O } < v \text{ O S}$$

and a ranking on the semantics||:

$$s^{\text{given}} \text{ O }^{\text{new}} < s^{\text{given}} \text{ O }^{\text{given}} < s^{\text{new}} \text{ O }^{\text{new}} < s^{\text{new}} \text{ O }^{\text{given}}$$

Do we predict the same? Almost. The four possible meanings are ordered in the indicated way if we assume that given is less marked than new. By frequency, the subject is given and the object new. One would expect that the frequencies prefer $s_{\text{nom}} \text{ O } v$, followed by $\text{O } s_{\text{nom}} v$, which are in turn followed by $s_{\text{nom}} v \text{ O}$ and $\text{O } s_{\text{nom}} v$. And this does not match the facts in Korean.

English is probably a better example of the same phenomenon. From semantic markedness, we find exactly the same candidate meanings and as English does not have syntactic variation in word order or case marking in this case, we would predict that the more marked semantic interpretations will be realised by ungrammatical forms. But, in fact the single order $s v \text{ O}$ expresses all four interpretations. My explanation is that for a form to be blocked for a certain marked interpretation it must be invariably interpreted as the unmarked meaning. That is not so in English: there is just a preference for the unmarked interpretation, but that can be overridden by determiners, intonation, other marking (e.g. another) and most importantly by the context. After all whether something can be interpreted as given depends on whether the referent is really given in the context. This has as a consequence that the unmarked form does not invariably mean the unmarked meaning and therefore that the marked meaning is not blocked for the unmarked form.

The development of other word orders is not necessary and counterbalanced by the need to mark subjects and objects by word order. If English has a word order constraint putting given before new, it is overridden by the constraint that puts the subject before the object. Korean is more liberal and can put given before new because of its case marking.

But that is not the complete explanation. If given comes before new, word order expresses givenness. It is therefore that Korean can —if case marking is present— express that the object is new and the subject is given by reversing the unmarked order. For a new subject and a new

|| I leave out one dimension of semantic variation in their example

object or for an old object and an old subject, there is no natural way of coding available.

Since old subjects and new objects are tendencies in natural corpora, the assumption of either subject before object or given before new, is a sufficient basis for having reinforcement of both subject before object and of given before new and therefore for the exploitation in language history of word order for either marking of grammatical relations or for marking given versus new. Without such a basis, there is no reason for expecting syntactic markedness to assume a semantic function.

It is however quite possible that longer chains form. Consider the pronouns: *me*, *himself*, *him*. But they seem to involve the existence of proper conventions (100% probabilities both ways) that *me* means first person object, and that *himself* means reflexivity before the chain effect occurs (*him* rules out first person and reflexivity, *himself* first person).

4. Constituents, Feature Spaces and Constraints

The theory of economy that I gave in the last section can be applied to nonconventional signs (e.g. the gestures that evolve in an attempt to communicate between you and somebody out of hearing on a raft in sea). Here there is no history and only the natural markedness orderings apply. But we find all three elements: the form, what is represented by the form and how the representation relates to the form. And blocking effects, though not with the same force as in conventional systems.

What comes into existence in a non-conventional communication is a mapping from parts of the sign to parts of its meaning (if the sign and the meaning are complex). We can take this as the basis for our signs.

We have to fix ideas here in order to make sense. The concrete decisions are not so important, except for the explanation. We divide DRSs into parts: the part that is given in the DRS representing the context and the part that is new to the sentence. The DR itself should represent the sentence and the resolutions of its presuppositions. Further divisions can and should be made, e.g. for distinguishing highly activated parts of the old material, but we will not pursue that here.

DRSs have discourse referents of various kinds and the discourse referents are related to each other by part whole relations, membership, thematic relations, ordering relations, etc. The association between syntax and semantics can be seen as identifying what parts of the sentence are concerned with the discourse referent in question. An association can be understood as a function from the discourse referents in the semantic representation to sets of words and morphs in the form.

This brings with it a notion of constituent: the image of a discourse referent under a mapping. But also the notion of an *argument*: a constituent that is part of a larger constituent is an argument of that larger constituent. It also gives a semantic version of syntactic relations in the thematic relations between the discourse referents to which they belong. And the old-new distinction among discourse referents gives a basic notion of information structure. Semantic sorts of discourse referents give a notion of classification for constituents, distinguishing nouns and verbs.

Crucially these notions are not the real thing when seen from a linguistic perspective. Constituents can here live below the word-level, they can be interrupted by intervening material and they may even fail to exist altogether. Linguistic constituents on the other hand can fail to be constituents in our sense. They may have a different classification and of course the semantic relations between the discourse referents of the constituents do not correspond directly with syntactic relations. Arguments also can appear to be away from their heads.

But our constituents, semantic relations, semantic categories can be regarded as the basis from which the notions that we know from linguistics have developed by processes

of language change. A syntactic role is an evolved semantic role. A category an evolved semantic sort and a morph an evolved word. A constituent is likewise a group of words referring to the same discourse referent that has developed coherence.

The features necessary for defining the linguistic concepts cannot be regarded as natural features of the elements of the sign. Instead, it is necessary to assume that they coevolve with the language and become observable features of words and constituents because they happen to be the necessary basis of a substantial generalisation in the particular language. It follows that they are not universal features that can be found in all languages and that allow of an a priori definition in terms of a conceptualisation of the world. Female gender plays a role in the agreement system of many languages, but it is never quite the same. The similarities between what goes on in different languages and the similarity in its appearance in different languages must be explained by the common origin in a central conceptual distinction much older than the human race, and the uniformities in the evolution of human languages, due to the communality of the conditions under which human languages evolve.

A development of a formal theory of linguistic signs needs therefore to appeal to a language dependent feature space. The features themselves are to words, constituents and relations between constituents by lexical specification. They form a realm of pseudoproperties and pseudorelations that need to be distinguished —if they need to be distinguished at all— purely on the basis that they make it simpler to account for the particular natural language and for its learnability. Linguistic features are classifications of the linguistic utterances of the language by the users of the language.

The feature space for a particular language develops alongside with the constraints for the language and gives the language in which the constraints are expressed. I am assuming that the feature space as given in versions of GPSG and HPSG for English is roughly correct. We have categories and subcategorisation in terms of these categories, agreement features, wh-features and in addition a topic feature. We also assume some semantic features, e.g. definiteness, mass and negation. In addition, syntactic relations, possibly defined in terms of subcategorisation.

The feature space defines which constraints are possible in a language. We can limit ourselves to simple operations like \rightarrow (one feature combination entails the other), $<$ (one feature bundle occurs before the other), $*$ (the two feature combinations are disallowed) and operations like *subject* (the subject of a constituent) and *head* (the head of a constituent). It is always possible to distinguish heads and dependents in our view of constituent: a dependent is just a constituent that is part of another constituent. That makes it also possible to define heads: it is that part of the constituent that is not an argument and shares the category of the constituent.

Under favourable circumstances, like in English, we can therefore define a subject as the nominal argument of a verbal category that binds the highest thematic role that is bound in the semantics[¶]. And objects as the next higher one. And we can define singular, plural, first and second person in equally standard and simple ways.

Given a space of features containing all the ones that seem useful, we can do some sign-based linguistics. Some useful constraints are stated in (7).

[¶] Dowty's (1991) theory of subject and object assignment corresponds to a simple hierarchy of OT constraints

- (7) NP is uninterrupted
 S is uninterrupted
 VP is uninterrupted
 AP is uninterrupted
 PP is uninterrupted
 S's subject agrees with its head
 S's subject comes before its head
 S's object comes after its head
 WHs come first

Those are all constraints making syntactic forms marked. Lexical constraints will enforce subcategorisation but most importantly will restrict possible interpretations.

- (8) $horse \rightarrow horse(x)$
 $horse \rightarrow heroine(x)$
 $horse \rightarrow \neg human(x)$
 $walk \rightarrow walk(e, agent(e, x), subject : x)$
 $walk \rightarrow walk(x, path(x, y))$
 $walks \rightarrow walk(e, agent(e, x), subject : x, 3sg : x)$
 $sees \rightarrow see(e, experiencer(e, x), subject : x, 3sg : x, theme(e, y), object : y)$
 $sees \rightarrow see(e, experiencer(e, x), subject : x, 3sg : x, theme(e, y), object : y, activity(z), comp : z)$
 $heroine(X) \rightarrow horse$
 $horse(X) \rightarrow horse$

Some explanation: *subject:z* places a requirement on the mapping, the subject of *see* should refer to *x*. Likewise *3sg:x* requires the NP denoting *x* to be third singular.

The two last constraints show that for each constraint in one direction, there is another in the other direction. The strength of an arrow one way is independent of the arrow in the other direction. The situation may arise that an interpretation I is unavoidable for F, but that I is more likely expressed by G, or the other way round.

The last group of constraints are semantic markedness constraints. We can have things like **old* < **new* < **inconsistent* And **participant* < **topic* < **given* < **known* < **connected* < **new*, which each class corresponding to a constraint and a natural ordering obtaining between them.

All of the constraints mentioned can be ranked by the gradual learning algorithm and will in that case lead to grammars that reproduce the frequencies in the corpus. I favour Jäger (too appear)'s bidirectional approach: ranking takes place on the basis of two questions: am I understood with this utterance?, and, would I have said it in the same way? The GLA only uses the second question. The bidirectional algorithm brings in a functional motivation: in trying to be understood more frequently we may not reproduce the corpus from which we learned, and provides a way of conceiving of a functionally driven theory of language change. We assume that this leads to weights for each of the constraints involved which will come into play if there are conflicts in interpretation or generation. What I have presented in this section is a way in which a very simple conception of markedness —frequency— can be brought into contact with standard ideas about the existence of language. Frequency requires concepts of the material to be counted. But if we have these concepts (the feature space), we automatically have constraints: the best ways of formulating statistical dependencies in the feature space that cannot be reduced to other dependencies. So we have reinvented OT, as a theory of implementation for statistically based markedness and above all as a theory of language learning.

There is however a difference. Each constraint now has its proper place. Lexical constraints constrain the association between form and meaning, syntactic markedness and semantic markedness are separate systems. Are they independent? The answer is yes and no. A least marked semantic representation for a form is not its meaning, if there is less marked form that blocks it for that meaning. A least marked form for an interpretation is not its proper expression if there is a less marked interpretation available.

Apart from these blocking effects there is no interference, syntax and semantics go their own ways. Let us look at a typical constraint.

subject before object

Since the subject is often a theme or old, there is an independent reason why it is followed, if there exists another constraint that places the topic before the focus. In that case, the constraint will have some weight but not very much. As long as it is entirely clear what is the subject and the object (e.g. by strong case marking or by headmarking) transgressions of the constraint will never contribute to misunderstandings. This changes if there is no or less marking of the subject and the object. There is then functional pressure behind it and the rule gets more weight: the word order starts to be a way of marking subjects and objects. The marking is not by an interpretational rule or by a convention: it is just that interpretations of NP-NP as object-subject are going to be blocked by the fact that there is a less marked form, the subject-object form. An association constraint making the first NP the subject can arise since the necessary concepts are in the feature space, but is not necessary for getting the effect. If we have a rule that places the topic before the comment of the sentence there is a similar effect. Though it is a constraint on the form and not on the interpretation, topic NPs are typically given and—in the absence of markings of the contrary—the earlier NP will normally be interpreted as given. Again this can make it necessary to put NPs that have to be interpreted as new and are not marked for that outside the topic, i.e. after the NPs that are to be interpreted as given.

Effects of this kind give a functional explanation for obligatory and optional marking. An obligatory marker like *another* forcing a new interpretation on an NP when an old interpretation is possible and therefore favoured as the less marked one is a clear case. But also case marking systems and headmarking systems can arise as ways of eliminating misunderstandings in finding out which NP is the subject and which one the object as argued in Zeevat & Jäger (2002). Jäger (to appear) provides a simulation system of some aspects of the process that leads to it. The processes are historical and functional and intimately related to patterns of phonological decay that may remove useful ways of marking from the language and so give rise to new patterns of marking. All constraints that describe tendencies in language use play a role in interpretation and expression and the interaction between them is quite complex. But it does not follow that markedness in the expression and interpretation direction is due to exactly the same constraints. In fact, there is an argument against that. Using the same set of constraints in both directions is not the correct way to speak “as everybody else” or to interpret “as everybody else does”. This is achieved by particular constraints in both directions that could be trained by monodirectional learning algorithms (to the extent that they have to be learned). Bidirectional learning optimises a correct balance between the concerns of interpretation and expression and lets the interpretational system and the expressive system influence each other.

5. Conclusion

In a weighted constraints framework where weights are connected with frequencies that there are absolutely marked forms (weight 0 means never, weight 1 means always). This explanation is in some cases to be preferred over the bidirectional explanations of ineffability

and incomprehensibility. Pure syntactic incorrectness is then not the fact that there is a less marked expression for the meaning that it would express, but can also mean that violations of absolute markedness constraints are involved.

Deblocking as described in Gärtner (to appear) is unproblematically in this framework. A simple example is from Bresnan (2001) about the distribution of Chichewa pronouns. Bound forms normally express topicality of the antecedent, free forms are used when the antecedent is not topical. In our framework, this comes out as the unmarked form (the phonologically less complex bound form) is assigned to the unmarked topical meaning (the older and more familiar the less semantically marked) and the association of free pronouns with nontopics arises through pragmatics as in section 1. It is the rule that maps topical interpretations to bound forms which blocks the free pronoun for topics and blocks the topic interpretation for free pronouns. This blocking disappears in the cases where Chichewa lacks a bound form.

Acknowledgements This paper is an attempt to draw conclusions from discussions with Reinhard Blutner, Gerhard Jäger, Anna Pilatova, Hans-Martin Gärtner about the Horn example. I thank Reinhard Blutner for reading a draft version and the workshop reviewers for their comments.

6. References

- Blutner, R. (2000). Some aspects of optimality in natural language interpretation. *Journal of Semantics*, 17, 189-216.
- Bresnan, J. (2001). The emergence of the unmarked pronoun. In J. G. G. Legendre & S. Vikner (Eds.), *Optimality-Theoretic Syntax*, (pp. 113-142). Cambridge (MA): MIT Press.
- Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 67, 547-619.
- Gärtner, H.-M. (to appear). On the OT status of unambiguous encoding. In R. Blutner H. Zeevat (Eds.), *Pragmatics in Optimality Theory*. Palgrave.
- Grice, P. (1975). Logic and conversation. In P. Cole J. Morgan (Eds.), *Syntax and Semantics 3: Speech Acts*, (pp. 41-58). New York: Academic Press.
- Hayes, S. C. Hayes, L. (1989). The verbal action of the listener as the basis of rule governance. In S. C. Hayes (Ed.), *Rule governed behavior: Cognition, contingencies, and instructional control*, (pp. 153-190). New York: Plenum Press.
- Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicatures. In D. Schirin (Ed.), *Meaning, Form and Use in Context*, (pp. 11-42). Washington: Georgetown University Press.
- Jäger, G. (2002). Some notes on the formal properties of bidirectional optimality. *Journal of Logic, Language and Information*, 11, 427-451.
- Jäger, G. (to appear). Learning constraint subhierarchies. the bidirectional gradual learning algorithm. In R. Blutner H. Zeevat (Eds.), *Pragmatics in Optimality Theory*. Palgrave.
- Karttunen, L. (1976). Discourse referents. In J. McCawley (Ed.), *Syntax and Semantics (Volume 7)*. New York: Academic Press.
- Lee, H. Beaver, D. (to appear). Input-output mismatches in OT. In R. Blutner H. Zeevat (Eds.), *Pragmatics in Optimality Theory*. Palgrave.
- Levinson, S. (2000). *Presumptive Meanings. The Theory of Generalized Conversational Implicatures*. Cambridge: MIT Press.
- Smolensky, P. (1996). On the comprehension/production dilemma in child language. *Linguistic Inquiry*, 27, 720-731. Tesar, B.
- Smolensky, P. (2000). *Learnability in Optimality Theory*. Cambridge (MA): MIT Press.

Zeevat, H. (2001). The asymmetry of optimality theoretic syntax and semantics. *Journal of Semantics*, 11, 243-262.

Zeevat, H. Jaeger, G. (2002). A statistical reinterpretation of harmonic alignment. In D. de Jongh, M. Nilsenova, H. Zeevat (Eds.), *Proceedings of the 4th Tblisi Symposium on Logic, Language and Linguistics*. Amsterdam: ILLC, Amsterdam, ICLC, Tblisi.

Zipf, G. (1949). *Human behavior and the principle of least effort*. Cambridge: Addison-Wesley.