Optimality Theory
as a General Cognitive Architecture

Workshop at the 33rd annual meeting of the Cognitive Science Society

July 20, 2011 in Boston, Massachusetts

Website: http://www.birot.hu/events/OTGCA/

Date and time: July 20, 9:00-12:30.

Location: White Hall Room (of the Boston Park Plaza Hotel).

9:00 Introduction (Judit Gervain and Tamas Biro)

9:10 Keynote address by Paul Smolensky: *Parallel Distributed Symbol Processing: Well-formedness optimization and discretization in cognition*

9:55 Giorgio Magri: *A comparison between OT and HG from a computational perspective*

10:20 Poster session, followed by coffee break

10:50 Petra Hendriks: *Asymmetries between production and comprehension and the development of Theory of Mind*

11:15 Douglas M. Jones: *Linguistic grammar and moral grammar: The case of kinship*

11:40 Lotte Hogeweg: *Optimality Theory as a general linguistic theory*

12:05 Géraldine Legendre and Mary Schindler: *Bilingualism and the optimizing of code-switching*

12:30 Business meeting

Posters in the poster session:

Stephen Goldberg & Ariel Goldberg: *Constraint interaction in the inscription of Chinese characters*

Ann Irvine, Mark Dredze, Geraldine Legendre and Paul Smolensky: *Optimality Theory syntax learnability: An empirical exploration of the perceptron and GLA*

Richard Mansell: *Translation universals: Can Optimality Theory help?*

Nazarré Merchant: *Using the fusional closure to assist in learning ranking information*

Clàudia Pons-Moll: *From positional faithfulness to contextual markedness: Phylogenetic aspects of OT/HG approaches*

Igor Yanovich: *The Logic of OT rankings*

Tamás Biró: *Religious mental structures: Counterintuitiveness represented in Optimality Theory*

20:00 Meeting at the entrance of the hotel, and leaving together for dinner.

Abstracts

Paul Smolensky (Johns Hopkins University): *Parallel Distributed Symbol Processing: Well-formedness optimization and discretization in cognition*

Does Optimality Theory provide a satisfactory basis for cognitive modeling of on-line processing? The description of OT as a competence theory—in which potentially infinitely many candidates are each evaluated by all constraints and then compared—is often mistaken as a performance model. But it is of course a basic characteristic of computation theory that an efficient algorithm (a processing theory) rarely corresponds in any direct way to the most insightful characterization (a competence theory) of the function the algorithm computed by the algorithm. OT computation was originally derived from a connectionist-grounded cognitive architecture, and in this talk I will describe how (continuous) connectionist networks can compute (discrete) outputs of OT grammars. These network computations bear no connection whatever to the sequential evaluation, by a sequence of constraints, of an infinite sequence of symbolic candidates. In fact, the only symbolic candidate ever represented in the processing system is the final output. Examples will be given illustrating gradient performance effects resulting from OT competence grammars in phonological production and syntactic comprehension. Outside the domain of language, an application of OT to cross-cultural variation in moral systems—who sleeps with whom—will also be briefly described.

Petra Hendriks (University of Groningen, Netherlands): *Asymmetries between production and comprehension and the development of Theory of Mind*

Children's development of production and comprehension do not always go hand in hand. Particularly puzzling are cases where children's production seems to be well ahead of their comprehension, as with object pronouns in languages such as English and Dutch. Such asymmetries between production and comprehension present a challenge to rule-based theories of language, but receive a straightforward explanation in constraint-based frameworks like Optimality Theory. A crucial aspect of such an explanation is the assumption that the adult grammar is the result of bidirectional optimization, which formalizes the idea that mature listeners take into account the speaker's options and choices and vice versa. This talk will discuss the implications of this view for language acquisition and its relation to the development and use of Theory of Mind reasoning.

Doug Jones (University of Utah): *Linguistic grammar and moral grammar: The case of kinship*

What is the scope of Optimality Theory? At its narrowest, it might be a theory of phonology only. At its broadest, it might be a theory of cognition in general. I use human kinship as a test case to argue for an intermediate position. OT seems to account for variation in kin terminology, suggesting that it may be useful in explaining the "grammar" of some semantic fields. Beyond this, it may also be useful in accounting for typological variation in rules of marriage, which suggests a role for OT in "moral grammar."

Lotte Hogeweg (Radboud University Nijmegen, Netherlands): *Optimality Theory as a general linguistic theory*

Optimality Theory is mostly associated with phonology but it has been applied to practically all linguistic subfields, including syntax, semantics and pragmatics. In this talk I will discuss the differences that can be identified in the way OT is applied to several linguistic disciplines. In relation to this, I will address the application of OT to kinship terminology by Jones (2010) and discuss why the analysis has more in common with previous OT analyses in the domain of phonology than with OT analyses in the domain of (lexical) semantics.

Géraldine Legendre and Mary Schindler (Johns Hopkins University): *Bilingualism and the optimizing of code-switching*

When speaking to one another bilingual speakers routinely and unconsciously code-switch, i.e. switch from one language to another, producing 'mixed' speech. We discuss the case of Urban Wolof, a mixture of French and Wolof spoken in the cities of Senegal which challenges all previous attempts at characterizing code-switching in a constrained fashion. We argue that an Optimality-Theoretic approach straightforwardly captures its otherwise unexpected properties as well as predicts the range of code-switching patterns attested cross-linguistically.

A COMPARISON BETWEEN OT AND HG FROM A COMPUTATIONAL PERSPECTIVE

GIORGIO MAGRI

Optimality Theory (OT) and Harmonic Grammar (HG) differ because the former assumes a model of constraint interaction based on strict domination, while the latter assumes a weighted model of interaction. As Prince and Smolensky (1997) admit, "that strict domination governs grammatical constraint interaction is not currently explained". Yet, Legendre et al. (2006, 911-912) make two suggestions. The first suggestion is that OT's strict domination might have *algorithmic advantages*, in the sense that it "may enable quick-and-dirty optimization algorithms [...] to consistently find a single global [...] optimum, whereas arbitrarily weighted constraints typically lead such algorithms to produce widely varying solutions, each only a local optimum." The second suggestion is that OT's strict domination might have *learnability advantages*: "another possibility is that demands of learnability provide a pressure for strict domination among constraints", although they note that "it remains an open problem to formally characterize exactly what is essential about strict domination to guarantee efficient learning."

Both conjectures have been challenged in the recent literature. Pater (2009) compares OT and HG from an *algorithmic perspective*, and reaches the opposite conclusion. He advocates "the replacement of OT's ranked constraints with [HG's] weighted ones" based on the fact "that the resulting model of grammar [...] is compatible with well-understood algorithms for learning and other computations" and states that "the strengths of HG in this area are of considerable importance" (p. 1002). Furthermore, Riggle (2009) and Bane et al. (2010) compare OT and HG from a *learnability perspective*, using tools from Statistical Learning Theory. They show that the two frameworks pattern alike w.r.t. a classical measure of learning complexity, namely they have the same *Vapnik-Chervonenkis (VC) dimension*. And they thus conclude that, "though there may be factors that favor one model over the other, the complexity of learning [...] is not one of them". These recent papers show that computational phonology has entered a mature stage, characterized by stronger connections with the neighbouring field of Machine Learning.

Yet, In this talk, I challenge both of these recent conclusions, thus vindicating the initial conjecture of OT's optimality. I present a simple trick that allows algorithms for HG to be extended to OT. Thus, HG has no algorithmic advantages over OT, contrary to Pater's claim. Furthermore, I point out that the VC dimension is well known to be a rather course upper bound on learning-theoretic complexity (especially for the case of linear classifiers). And I build the case for a learnability advantage of OT over HG, based on some recent results in Koltchinskii et al. (2003b), Koltchinskii et al. (2003a) and Koltchinskii and Panchenko (2005).

References

Bane, Max, Jason Riggle, and Morgan Sonderegger. 2010. "The VC dimension of constraint-based grammars". *Lingua* 120.5:1194–1208.

Koltchinskii, Vladimir, and Dmitry Panchenko. 2005. "Complexities of convex combinations and bounding the generalization error in classification". *Ann. Statist.* 33.4:1455–1496.

Koltchinskii, Vladimir, Dmitry Panchenko, and Savina Andonova. 2003a. “Generalization bounds for voting classifiers based on sparsity and clustering”. In *Lecture Notes in Artificial Intelligence 2777*, 492–505.

Koltchinskii, Vladimir, Dmitry Panchenko, and Lozano. 2003b. “Bounding the generalization error of convex combinations of classifiers: Balancing the dimensionality and the margins”. *Ann. Appl. Probab.* 13:213–252.

Legendre, Geraldine, Antonella Sorace, and Paul Smolensky. 2006. “The Optimality Theory/Harmonic Grammar connection”. In *The Harminic Mind*, ed. Paul Smolensky and Geraldine Legendre, 903–966. Cambridge, MA: MIT Press.

Pater, Joe. 2009. “Weighted Constraints in Generative Linguistics”. *Cognitive Science* 33:999–1035.

Prince, Alan, and Paul Smolensky. 1997. “Optimality: From Neural Networks to Universal Grammar”. *Science* 275:1604–1610.

Riggle, Jason. 2009. “The Complexity of Ranking Hypotheses in Optimality Theory”. *Computational Linguistics* 35(1):47–59.

Constraint interaction in the inscription of Chinese characters

Stephen J. Goldberg[1] and Ariel M. Goldberg[2]
[1]Department of Art History, Hamilton College
[2]Department of Psychology, Tufts University

Chinese writing has a long and venerated tradition dating back over 3,000 years and serves a variety of purposes from the purely practical to the purely aesthetic. In the ordinary writing of Chinese characters (*xiezi* 寫字), the primary function is to convey the lexical meaning of the written character. In the art of Chinese calligraphic inscriptions (*shufa* 書法), however, the visual form of the written characters is simultaneously a means of communication and aesthetic visual expression. In this vein, the logographic nature of the Chinese orthographic system makes the characters particularly suitable to creative visual elaboration in the inscription of a text.

We propose to explore the possible application of Optimality Theory (OT) as a framework for the analysis of Chinese calligraphic inscriptions. Just as the figurative language of poetry is "a departure from the literal" (Bloom, 2004), the inscription of the character in *shufa* is a departure from the legible, so that its form is both "expressive and evocative." Our goal is to develop a theory of the way that violable *markedness* constraints enforce representational well-formedness in the structuring of calligraphic inscriptions. We aim therefore to extend the results of modern research on the role of constraints in phonological grammar to the field of Chinese calligraphy.

There are two possible domains of analysis. The first and most basic domain is to formally describe the characteristics of different calligraphic scripts (e.g., seal, clerical, standard, semi-cursive, and cursive script). The second domain concerns aesthetic stylistic norms specific to the art of calligraphic inscriptions. In this domain, the well-formedness of a calligraphic inscription lies in its accordance with period-specific stylistic norms that typically are a violation of previously established normative constraints. In the present work, we focus on the former domain: the writing of inscriptions in specific calligraphic scripts.

We propose that the formal properties of different calligraphic scripts can be understood as arising through the interaction of a set of violable constraints. We argue that coherence-generating 'markedness' constraints enforce visual well-formedness at three levels of organization: 1) the overall compositional arrangement of the inscription (e.g., the organization of characters within each column and between each row; *zhangfa* 章法), 2) the compositional structure of individual character formations (*jiegou* 結構), and 3) the forms of the individual brushstrokes (*bifa* 筆法). At the same time, a set of 'faithfulness' constraints ensures that the surface form of written characters is legible and resembles the canonical form enough to be recognized. We propose that the same set of constraints is operative across scripts and that what differentiates calligraphic scripts is the specific ranking of constraints.

In putting forth this theory, the ultimate goal is to make the shift to a cognitive perspective in the field of art history by investigating the explanatory force of a constraint-based analysis of the surface forms of Chinese calligraphic inscriptions.

Optimality Theory Syntax Learnability:
An Empirical Exploration of the Perceptron and GLA

Ann Irvine, Mark Dredze, Geraldine Legendre, and Paul Smolensky
Johns Hopkins University

This work brings together several threads of research on Optimality Theory (OT) and Harmonic Grammar (HG) learnability. As noted in previous work, including Pater (2008) and Magri (2010), the perceptron learning algorithm is well-established in the Machine Learning field and is a natural choice for modeling human grammar acquisition. The algorithm learns from one observation at a time, and it is capable of learning from a noisy corpus of observed natural language. In this work, we use the perceptron algorithm to learn a model that specifies a set of constraint weights relevant to one syntax phenomenon, Czech word order. We extract training data (sentences annotated with grammatical and information structure and their surface word orders) from the Prague Dependency Treebank (Hajic et al., 2001) and use basic alignment (edge-most) constraints on grammatical and information structure to predict the surface order of the subject, verb, and object. The perceptron algorithm learns a set of numeric, weighted constraints (a Harmonic Grammar). Ordering the constraints by the magnitude of their weights may specify a hierarchical constraint ranking (an OT Grammar), which is the essence of the classic Gradual Learning Algorithm (GLA) (Boersma, 1997). We describe and compare the two learning algorithms in detail and use a held out set of empirical data to quantitatively evaluate each. We show that by allowing for so-called *ganging-up-effects*, the more expressive Harmonic Grammar models Czech Word Order more accurately than the GLA OT grammar. Finally, crucially, it is also capable of modeling variation in production.

References

Boersma, P. (1997). How we learn variation, optionality, and probability. In *Proceedings of the Institute of Phonetic Studies*, Volume 21, pp. 43–58.

Hajic, J., B. Hladka, and P. Pajas (2001). The prague dependency treebank: Annotation structure and support. In *In Proc. of the IRCS Workshop on Linguistic Databases*, pp. 105–114.

Magri, G. (2010). Hg has no computational advantages over ot: consequences for the theory of ot online algorithms. *Manuscript, IJN, ENS. Available as ROA-1104*.

Pater, J. (2008). Gradual learning and convergence. *Linguistic Inquiry 39*(2), 334–345.

Translation Universals: Can Optimality Theory Help?
Richard Mansell, University of Exeter

This paper will analyse findings from the study of the translation process within an Optimality-Theoretic framework to determine how it can help us to understand cognitive aspects of translation.

There have been three attempts to apply Optimality Theory within translation studies; Dols and Mansell (Dols 2006; Mansell 2004, 2008; Dols and Mansell 2008), Darwish (2008) and Calfoglou (2010). It seems that these have arisen independently, since there is no cross-referencing between them, something that echoes this workshop organisers' general comment on various OT applications: 'The authors of these isolated attempts usually even did not know of each other.' These applications will be analysed to determine their strengths, but also areas where OT principles have not been fully applied, such as the identification of inputs and outputs, and the universality of constraints.

In particular I shall focus on the search for universals in translation, which has been a particularly fertile line of investigation in translation studies since the advent of corpus-based studies in the mid 1990s (Mauranen and Kujamaki 2004; see Baker 1993 for the seminal work in the field). Despite this activity, to date there is still no satisfactory definition of translation universals themselves, nor how they manifest themselves and how they should be studied. I propose that the definition of the object of study of universals has been determined to a great extent by the method of investigation (see Pym 2008 for a criticism of corpus-based studies in this field), and shall use optimality-theoretic principles to demonstrate provide an encompassing and yet clearly delimited definition of what translation universals are and how and why they arise, regardless of the language pair involved.

References

Baker, Mona. 1993. 'Corpus Linguistics and Translation Studies: Implications and

Applications' in Mona Baker et al. (eds), *Text and Technology*. Philadelphia and

Amsterdam: John Benjamins. 233-250.

Calfoglou, Christine. 2010. 'An Optimality Approach to the Translation of Poetry' in Antoinette Fawcett, Karla Guadarrama Garcia and Rebecca Hyde Parker (eds). *Translation: Theory and Practice in Dialogue*. London: Continuum. 85-105.
Darwish, Ali. 2008. *Optimality in Translation.* Melbourne: Writescope.
Dols, Nicolau. 2006. 'La teoria de l'optimitat aplicada a les traduccions. Dos exemples a partir d'*Os Lusiadas*' in *Actes del Primer Col.loqui de l'Associaci ó de Lusitanistes de l'Estat Espanyol.* Palma: UIB. 148-159
----- and Richard Mansell. 2008. 'Resolving Meaning Conflict in Translation'. *Linguistica Antverpiensia New Series*, 7, 45-74.
Mansell, Richard. 2004. 'Optimality Theory Applied to the Analysis of Verse Translation'. Available at http://roa.rutgers.edu.
-----. 2008. 'Optimality in Translation', in Anthony Pym and Alexander Perekrestenko. *Translation Research Projects I*. Tarragona: Intercultural Studies Group. 3-12.
Mauranen, Anna and Pekka Kujamaki (eds). 2004. *Translation Universals: Do They Exist?* Amsterdam and Philadelphia: John Benjamins.
Pym, Anthony. 2008. 'On Toury's Laws of How Translators Translate', in Anthony Pym et al. (eds). *Beyond Descriptive Translation Studies*. Amsterdam and Philadelphia: John Benjamins. 311-328.

Nazarré Merchant

Title: Using the fusional closure to assist in learning ranking information

A well-known learnability problem both within an OT/HG framework and without is that a phonological learner must construct a lexicon that matches observed surface forms and must produce a grammar that maps those constructed lexical items to those observed surface forms. Adopting Optimality Theory narrows the problem: grammatical learning consists solely of ranking determination. Even with this narrowing, the learner must produce a lexicon concomitantly with a ranking. The logic of Optimality Theory has allowed researchers to begin to address the simultaneous nature of the task by positing a number of differing techniques for extracting ranking information from unspecified and underspecified forms (Merchant and Tesar 2006, Jarosz 2006, Apoussidou 2007, amongst others). These techniques are known to fail in a number of situations stemming from a number of causes, often related to the form of the information generated from the learning algorithms. For example in CPR (Merchant 2008) ranking information is produced using MRCD from not fully specified forms producing sets of ERCs that are missing entailed and crucial ERCs, gaps in the set, that cause the algorithm to fail to capture relevant ranking arguments.

In this poster I present a means of filling these gaps by, for a set of ERCs, producing the set's *fusional closure*. The fusional closure of a set is the *fusional closure* of the set in the mathematical sense: it is that set that is *closed* under the operation of fusion (Prince 2002), so that the result of fusing any two ERCs in the set is also in the set. This set is useful for learning for a number of reasons: a set and its fusional closure are mutually entailing, allowing the learner to compute over fusional closures instead of base ERC sets. Furthermore, the fusional closure of a set of ERCs has the property that every ERC entailed by the original set is entailed by exactly one ERC in its fusional closure.

This final property of individual entailment is shown to be necessary for extracting ranking information from sets of ERCs using an algorithm such as CPR. CPR works by generating sets of ERCs using MRCD, one for each possible lexicon for a pair of overt forms and determining the shared ranking information across these sets of ERCs using the join (Merchant 2011) to extract shared information. The algorithm, computing over MRCD produced sets, misses ranking information because there are entailed ERCs that are not entailed by individual ERCs in the ERC sets, but by combinations of ERCs. Computing over fusional closures shores up this lack of individual entailment.

The ideas presented here emphasize that the logic of the linguistic system plays a crucial role in learning. Determination of ranking information from lexical items occurs because there is an entailment relationship between sets of ERCs consistent with given lexical hypotheses. Gaps in the information produced can be filled using computational tools built up from the linguistic system itself. Open questions still optain: how are these computational tools manifested in an ICS architecture and what constraints does an ICS system impose on learning theory? Even with these questions it is clear that the logic of the linguistic system can play a crucial role in linguistic, and possibly other types, of learning.

Keywords: Phonology, Learnability, ICS

References:

Apoussidou, Diana. 2007. The Learnability of Metrical Phonology. PhD. dissertation, University of Amsterdam, Amsterdam.

Jarosz, Gaja. 2006. Rich Lexicons and Restrictive Grammars - Maximum Likelihood Learning in Optimality Theory. PhD. dissertation, The Johns Hopkins University, Baltimore, MD. ROA-884.

Merchant, Nazarré. 2008. Discovering underlying forms: Contrast pairs and ranking. PhD. dissertation, Rutgers University, New Brunswick. ROA-964.

Merchant, Nazarré, and Tesar, Bruce. 2008. Learning underlying forms by searching restricted lexical subspaces. In *Proceedings of the Forty-First Conference of the Chicago Linguistics Society (2005), vol. II: The Panels*, 33-47. ROA-811.

Merchant, Nazarré, 2011. Learning ranking information from unspecified overt forms using the join. To appear in: *The Proceedings of the Forty-Seventh Conference of the Chicago Linguistics Society*.

Prince, Alan, 2002. Entailed Ranking Arguments. ROA-500.

Clàudia Pons-Moll

From positional faithfulness to contextual markedness

1. Introduction. Optimality Theory has commonly made use of two types of constraints: faithfulness and markedness constraints (Prince & Smolensky 1993, McCarthy & Prince 1995). Moreover, both kinds of constraints admit to be relativized according to the position or the context to which they apply or are active. Indeed, in addition to standard faithfulness constraints and context-free markedness constraints, *positional* faithfulness constraints (Beckman 1998; Casali 1996, 1997) and *contextual* markedness constraints (Prince & Smolensky 1993) are generally invoked. Both kinds of constraints, however, are alleged to be redundant and thus mutually excluding, in that they do the same job. The former just interact with context-free markedness constraints: the effects of a general markedness constraint can be inhibited by the higher ranking of a faithfulness constraint which protects a segment, a feature, etc., in a specific structural position. The latter, on the other hand, interact with standard faithfulness constraints: the effects of a specific markedness constraint can be reduced by relativizing it to a specific context.

2. Goal. The purpose of this paper is to explore the relation between positional faithfulness and contextual markedness constraints, and to show how, in some particular cases and from a diachronic point of view, the latter can be interpreted as induced by the effects of the former into the grammar of languages throughout its historical development. Our proposal is illustrated with a set of cases of underapplication of vowel reduction which occur in some dialects of Catalan.

3. Data. In Majorcan Catalan (MC), the process of vowel reduction of the mid front vowels /e/ and /ɛ/ to schwa [ə] in unstressed position underapplies under certain circumstances: *a*) in productive derived forms with an unstressed vowel located in the initial syllable of the stem, which alternates with a stressed mid front vowel in the stem of the underived form (*p*[é]*ix* 'fish' ~ *p*[e]*ixet* 'fish *dim.*'; see also (1)); *b*) in verbal forms with an unstressed vowel located in the initial syllable of the stem, which alternates with a stressed close mid front vowel in another verbal form of the same inflectional paradigm (*p*[é]*ga* '(s/he) hits' ~ *p*[e]*gam* '(we) hit'; see also (2)); *c*) in learned and loan words with an unstressed *e* located in the initial syllable of the stem (*p*[e]*culiar* 'peculiar'; see also (3)).

4. Proposal. 4.1. *Alternating forms* (cases *a* and *b*). In Pons-Moll (in press a, b), it is argued that underapplication of vowel reduction to schwa in MC derivational and inflectional forms is a direct consequence of the interaction of the prominence constraint hierarchy banning certain vowels in unstressed position according to their sonority value and a set of output to output faithfulness constraints relativized according to two factors: the productivity of the derivational process and the position of the affected vowel within the stem (see the referred works for a more formal details about the proposal). 4.2. *Non-alternating forms* (cases *c*). Underapplication of vowel reduction to schwa in learned and loan words is also circumscribed to those cases in which the unstressed vowel is located in the initial syllable of the stem. In these cases, however, the unstressed vowel does not alternate with a stressed one. O-O *positional faithfulness constraints*, therefore, cannot explain this behavior, but *contextual markedness constraints* banning a schwa in this specific position (*i.e.*, the initial syllable of the stem) can. From a diachronic perspective, a plausible explanation of these facts is to consider that the activity of the O-O faithfulness constraints relativized according to the position of the vowel within the stem, responsible for underapplication of vowel reduction to schwa in productive derivation and inflection, that is, in the productive phonology of the dialect, and which have provoked a drastic reduction of the occurrences of the schwa in stem-initial position, have led, throughout time, to a reinterpretation of the unstressed vowel system by MC speakers. That is to say, the effects of the positional faithfulness constraints enhancing the appearance of [e], instead of [ə], in the initial syllable of the stem, would have been reinterpreted by MC speakers as a consequence of a contextual or positional markedness constraint of the type *ə/Initial-Syll-Stem, banning a schwa in the initial syllable of the stem and which at present is operating just in loanwords. Furthermore, the prediction is that this constraint will likely affect all kinds of words, motivating the massive disappearance of the schwa in this specific position.

DATA AND REFERENCES

(1) *Normal application* vs. *underapplication of vowel reduction in derivation*

BASE (UNDERIVED FORM)		PRODUCTIVE DERIVATION		NON-PRODUCTIVE DERIVATION	
a. Stressed stem with [é] *or* [ɛ́]		*b. Unstressed stem with the vowel in the initial syllable of the stem* → unexpected [e]		*c. Unstressed stem with the vowel in the initial syllable of the stem* → expected [ə]	
p[é]*ix*	'fish'	*p*[e]*ix*[ə́]*t*	'fish *dim.*'	*p*[ə]*ixat*[é]*r*	'fisherman'
t[ɛ́]*rra*	'earth'	*t*[e]*rr*[ə́]*ta*	'earth *dim.*'	*t*[ə]*rr*[é]*stre*	'terrestrial'
d. Stressed stem with [é] *or* [ɛ́]		*e. Unstressed stem with the vowel* not *in the initial syllable of the stem* → expected [ə]		*f. Unstressed stem with the vowel* not *in the initial syllable of the stem* → expected [ə]	
pap[é]*r*	'paper'	*pap*[ə]*r*[ə́]*t*	'paper *dim.*'	*pap*[ə]*rera*	'paper basket'
fid[ɛ́]*u*	'noodle'	*fid*[ə]*u*[ə́]*t*	'noodle *dim.*'	*fid*[ə]*u*[á]*da*	'noodle dish'

(2) *Normal application* vs. *underapplication of vowel reduction in inflection*

STRESSED-STEM VERBAL FORM	UNSTRESSED-STEM VERBAL FORM
a. Stressed stem with [é] or [ɛ́]	*b. Unstressed stem with the vowel in the initial syllable of the stem* → unexpected [e]
p[é]*ga*, *p*[é]*gues*, *p*[é]*gui*, *p*[é]*guis*, *p*[é]*guen* 'to hit' verbal forms	*p*[e]*g*[á]*m*, *p*[e]*g*[á]*u*, *p*[e]*gar*[é], *p*[e]*gar*[í]*es*... 'to hit' verbal forms
esp[é]*r, esp*[é]*res, esp*[é]*ra, esp*[é]*ri*, *esp*[é]*rin* 'to wait' verbal forms	*esp*[e]*r*[á]*m*, *esp*[e]*r*[á]*u*, *esp*[e]*r*[á]*ssis* 'to wait' verbal forms
c. Stressed stem with [ɛ́]	*d. Unstressed stem with the vowel in the initial syllable of the stem* → expected [ə]
x[ɛ́]*rr*, *x*[ɛ́]*rra*, *x*[ɛ́]*rren*, *x*[ɛ́]*rris*, *x*[ɛ́]*rren* 'to chat' verbal forms	*x*[ə]*rr*[á]*m*, *x*[ə]*rr*[á]*u*, *x*[ə]*rrar*[í]*es* 'to chat' verbal forms
at[ɛ́]*rra*, *at*[ɛ́]*rren*, *at*[ɛ́]*rri*, *at*[ɛ́]*rrin* 'to land' verbal forms	*at*[ə]*rr*[á]*m*, *at*[ə]*rr*[á]*u*, *at*[ə]*rrar*[í]*es*... 'to land' verbal forms
e. Stressed stem with [é]	*f. Unstressed stem with the vowel* not *in the initial syllable of the stem* → expected [ə]
cont[é]*st*, *cont*[é]*stes*, *cont*[é]*sta*... 'to answer' verbal forms	*cont*[ə]*st*[á]*m*, *cont*[ə]*st*[á]*u*, *cont*[ə]*star*[í]*a*... 'to answer' verbal forms
acc[é]*pt*, *acc*[é]*ptes*, *acc*[é]*pta*... 'to accept' verbal forms	*acc*[ə]*pt*[á]*m*, *acc*[ə]*pt*[á]*u*, *acc*[ə]*ptar*[í]*a*... 'to accept' verbal forms

(3) *Normal application* vs. *underapplication of vowel reduction in inherited and loanwords*

a. LEARNED AND LOAN WORDS		*b*. INHERITED WORDS	
p[e]*culi*[á]*r*	'peculiar'	*p*[ə]*ssig*[á]*r*	'to pinch'
p[e]*d*[á]*l*	'pedal'	*b*[ə]*s*[á]*da*	'kiss'
p[e]*l·l*[í]*cula*	'film'	*b*[ə]*ss*[ó]	'twin'
comm[e]*mor*[á]*r*	'to commemorate'	*m*[ə]*nt*[í]*da*	'lie'
llargm[e]*tr*[á]*tge*	'feature film'	*m*[ə]*l*[ó]	'melon'
imp[e]*c*[á]*ble*	'impeccable'	*p*[ə]*ned*[í]*r-se*	'to regret'
m[e]*dic*[í]*na*	'medicine'	*m*[ə]*norqu*[í]	'Minorcan'
f[e]*l*[í]*ç*	'happy'	*f*[ə]*ix*[ú]*c*	'heavy'
f[e]*titx*[í]*sme*	'fetishism'	*f*[ə]*r*[í]*r*	'to hurt'
v[e]*rm*[ú]*t*	'vermouth'	*v*[ə]*ll*[ú]*t*	'velvet'
v[e]*rb*[é]*na*	'party'	*v*[ə][í]	'neighbor'

Selected references: Pons-Moll, Clàudia (accepted, in press) «Underapplication of vowel reduction to schwa in Majorcan Catalan productive derivation, verbal inflection and learned and loan words». A: *Proceedings of Going Romance 2009*. Amsterdam-Philadelphia: John Benjamins. Clàudia Pons-Moll (in press) «Underapplication of vowel reduction to schwa in Majorcan Catalan. Some evidence for the left syllable of the stem as a prominent position and for subparadigms». A: Kan, S.; Moore-Cantwell, C.; Staubs, R. (ed.) *Proceedings of NELS 40*. Prince, A.; Smolensky, P. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. New Brunswick / Boulder: Rutgers University / University of Colorado. Published (2004), Malden / Oxford: Blackwell.

THE LOGIC OF OT RANKINGS

IGOR YANOVICH

Standard OT learning algorithms such as RCD or GLA use a single total ordering or partial ordering (stratified hierarchy) of constraints as the hypothesis about the grammar. As is well known, this necessarily leads to overcommitment, because certain datasets (think of a comparative row like [W, W, L]) impose irreducibly disjunctive conditions on what a ranking compatible with the data should be. As the result, learning is never conservative, and hypotheses often have to be rolled back even when the data are perfectly consistent.

At the same time, it is clear that we could in principle use the set of all total rankings compatible with the data as a conservative hypothesis, but no method for working with such sets was developed other than brute force ones. And those are both computationally heavy, and unintuitive: it is not easy for humans to work with amorphous sets of total rankings. As the result, even most basic facts about sets of rankings which constitute faithful, non-overcommiting grammar hypotheses for tableaux, were not known: for instance, which sets of rankings can be such hypotheses, and which cannot correspond to any tableau at all.

This paper overcomes the technical problems of working with sets of total rankings, and develops methods which allow to use them as full-fledged grammar hypotheses, and thus enable learning without overcommitment. There are two key components of our approach:

- We conservatively extend OT compatibility conditions from total rankings to partial rankings and sets of partial rankings. A set of partial rankings comes out as OT-equivalent to a set of all of its refinements, and can be used in its place in reasoning, which makes such reasoning both more intuitively accessible and computationally easier.
- We find a one-one correspondence between (equivalence classes of) OT tableaux and (equivalence classes of) sets of OT rankings. The bijection is computable, and thus provides a way to transform the data in the form of a tableau without any overcommitment into the set of rankings containing all and only the information which was present in the data.

Thus OT tableaux and sets of rankings turn out to be two sides of the same coin. Thus there is no actual choice between using (objects based on) rankings or (objects based on) tableaux (cf. the position of [Prince, 2010]) as grammar hypotheses: they are just different ways of expressing the same information.

Methods for manipulating tableaux and extracting information from them about the ordering of certain constraints are well-known. As an addition to them, the paper develops methods for manipulating rankings as well. For instance, we define an operation of pairwise ranking-union $\Cup$ on sets of partial rankings which corresponds to the union of tableaux, and equivalence-preserving transformations on sets of rankings simplifying them, which constitutes an analogue on the tableau side operations used in algorithms such as Fusional Reduction of [Brasoveanu and Prince, 2005].

References

[Brasoveanu and Prince, 2005] Brasoveanu, A. and Prince, A. (2005). Ranking and necessity, part i. ROA 794.
[Prince, 2010] Prince, A. (2010). Ot without ranking. Handout for Harvard Linguistics Colloquium, February 2010.