Frans Adriaans, Natalie Boll-Avetisyan, & René Kager
{frans.adriaans, natalie.boll-avetisyan, rene.kager}@let.uu.nl
Utrecht Institute of Linguistics OTS, Utrecht University, The Netherlands

Abstract Phonotactic Constraints for Speech Segmentation: Evidence from Human and Computational Learners

Mittwoch, 04.03.2009, 15.00 – 15.30

Studies of phonological systems support the view that phonotactic constraints are abstract in the sense of referring to natural classes. An example is OCP-Place (McCarthy, 1988). In many languages, this constraint has the gradient effect that labial-labial pairs across a vowel are underrepresented in the lexicons (Arabic: Frisch, Pierrehumbert & Broe 2004; Dutch: Kager & Shatzman, 2007). The present paper argues that abstract phonotactic constraints are both psychologically real and computationally learnable. We provide converging evidence for this hypothesis from both psycholinguistic and computational learning experiments.

We address the psychological reality of abstract phonotactic constraints by examining their potential role in speech segmentation. Although it is widely accepted that the segmentation of continuous speech into discrete, word-sized units is aided by phonotactics (e.g., McQueen, 1998), the role of abstract phonotactic constraints in segmentation has not been addressed. We argue that speech segmentation is aided by abstract phonotactic constraints that are the result of generalization over segment co-occurrence distributions.

A series of artificial language learning experiments tested whether human learners have indeed internalized an abstract constraint, OCP-Place, and use this constraint as a cue for segmentation. The artificial languages to which Dutch participants were exposed consisted of a stream of CV syllables starting with labials (P) and coronals (T) (e.g. … TPPTPPT...). As predicted, learners used OCP-Place, resulting in a significant preference for PTP (which respects the constraint) over PPT and TPP segmentations (which violate it).

Our computational model aims at pinpointing the mechanisms that allow humans learners to construct abstract constraints from input data. Previous simulations with our model showed that abstracting over statistically learned biphone constraints improves the segmentation performance of the learner. Input to the learner consisted of unsegmented utterances from the Spoken Dutch Corpus. After an initial statistical analysis of biphones, the model induces constraints, which are then generalized into abstract constraints by feature-based generalization. A series of computer simulations showed that the combination of statistical learning and feature-based generalization outperforms purely statistical learning, thereby indicating a potential role for abstract constraints in human segmentation.

We are currently matching predictions of the computational model with human segmentation performance in the behavioral experiments. By training the model on Dutch input, testing it on segmentation of the artificial language, and relating its segmentation output to human segmentation performance, we aim at providing a unified account of the psychological reality and computational learnability of abstract phonotactic constraints.

References:

Frisch, S., Pierrehumbert, J., & Broe, M. (2004). Similarity avoidance and the OCP. *Natural Language and Linguistic Theory,* **22**, 179-228.

Kager, R., & Shatzman, K. (2007). Phonological constraints in speech processing. In B. Los & M. van Koppen (eds.), *Linguistics in the Netherlands 2007*. Amsterdam: Benjamins.

McCarthy, J. (1988). Feature geometry and dependency: A review. *Phonetica,* **43**, 84–108.

McQueen, J. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language,* **39**, 21-46.

Diana Apoussidou
dapou1@linguist.umass.edu
University of Massachusetts Amherst/University of Amsterdam

Learning allomorphy

Mittwoch, 04.03.2009, 18.00 – 18.30

This talk gives an account for how allomorphy can be learned at the same time as and with the help of phonological grammar. Previous works on the learnability of underlying representations in Optimality Theory have either given accounts on e.g. how grammar and lexicon can be acquired in turns (for instance, Tesar et al. 2003) or by calculating probabilities (Jarosz 2006), or by incorporating lexical constraints into the grammar (Apoussidou 2007). The learnability of allomorphy is a special case of learning underlying representations, because the learner needs to not only acquire the underlying representation of a form, but furthermore needs to acquire different phonological underlying representations for a lexical entry depending on a given phonological context. In the case of the indefinite article in English, the decision between the two allomorphs *a* and *an* depends on whether the following word starts with a vowel or a consonant: *an apple* but *a pen*. Children acquiring English therefore need to learn the two underlying representations for the indefinite article, and they need to learn the phonological grammar to determine what allomorph to use in which context. This was modelled with computer simulations using a basic online learning algorithm for underlying forms (Apoussidou 2007), where lexical representations are encoded as constraints. In these simulations, the virtual learners acquired the appropriate allomorphs of a given word by crucially relying not only on occurrences of forms such as *an apple* or *a pen*, but also on words that contained clues about the phonology of the language without displaying variation, i.e. forms that aided the acquisition of the phonology only, and not the acquisition of the lexicon. This approach of lexicon acquisition as computation (i.e. constraint ranking) enables direct interaction of the lexicon with the grammar of a language. Acquiring a lexicon can proceed as acquiring a grammar, and no specific learning mechanisms for acquiring a lexicon are needed.

References

Apoussidou, D. (2007). The learnability of metrical phonology. Dissertation, LOT Dissertation Series 148.

Jarosz, G. (2006). Rich lexicons and restrictive grammars - Maximum likelihood learning in Optimality Theory. Dissertation, Baltimore, Maryland.

Tesar, B., J. Alderete, G. Horwood, N. Merchant, K. Nishitani & A. Prince (2003). Surgery in language learning. In *Proceedings of the 22nd West Coast Conference on Formal Linguistics* (WCCFL), eds. G. Garding and M. Tsujimura, 477-490. Somerville, MA: Cascadilla Press. ROA 619.

Tamás Biró
birot@nytud.hu
University of Groningen (RUG) and University of Amsterdam (UVA)

Performance of Learning & Learning from Performance

Mittwoch, 04.03.2009, 16.30 – 17.00

Formal models of language learning suppose that the learning data are produced by the target grammar. In other words, the learner supposedly develops her linguistic competence by having direct access to the product of the linguistic competence of the teacher. Some algorithms are robust to reasonable random noise, but are still based on the idea that most of the learning data reflect the target grammar.

In reality, however, the learner is exposed to the linguistic *performance* of the teacher, not to his *competence*. As long as performance errors are seen as random noise on competence, a view going back to Chomsky's introduction of the competence-performance distinction in his *Aspects*, the robust algorithms just referred to may be sufficient. Yet, performance can be argued to be more complex than mere noise. Both Bíró (2006) and Smolensky and Legendre (2006) argue that linguistic performance can be modelled as the algorithm that implements computationally the function describing competence. In particular, they suggest performance errors are locally optimal candidates of an Optimality Theoretical candidate sets, which can trap the performance algorithm, simulated annealing. Bíró (2006) also argues that in certain cases performance error forms are produced with a significant percentage even if simulated annealing is given ample computing time.

Consequently, this talk will argue, research on the learnability of linguistic models must take into account the divergences between competence and performance. The paper will first present how standard learning algorithms in Optimality Theory must be revised to meet the proposed switch in the research paradigm (Biró 2007). Second, results of a few simple experiments will be reported, which can be evaluated against child language phenomena. Finally, we draw the consequences of the approach for language evolution models based on iterative learning.

References

Tamás Bíró. F*inding the Right Words: Implementing Optimality Theory with Simulated Annealing*. PhD thesis, University of Groningen, 2006. ROA-896.

Tamás Biró. 'The benefits of errors: Learning an OT grammar with a structured candidate set'. In *Proceedings of the Workshop on Cognitive Aspects of Computational Language Acquisition*, pages 81–88, Prague, Czech Republic, June 2007.

Paul Smolensky and Géraldine Legendre (eds.). *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar.* MIT Press, Cambridge, 2006.

Paul Boersma
paul.boersma@uva.nl
University of Amsterdam

Category emergence in OT

Freitag, 06.03.2009, 12.30 – 13.30

Learning algorithms for Optimality Theory (Tesar 1995; Tesar & Smolensky 1993, 1998; Boersma 1997; Boersma & Hayes 2001; Boersma to appear) have been very good at ranking constraints, and can even moderately well handle cases with three representations (Tesar 1997; Tesar & Smolensky 2000; Boersma 2003, 2008; Apoussidou & Boersma 2004) or even four representations (Apoussidou 2007).

A prerequisite for these successes is that the phonological categories and the constraints that reference these categories are given to the learner in advance. This is not a natural situation in human language acquisition. Every language has a different set of categories, and these have to be created on the basis of language input (if you believe in an innate set of categories that is pruned during language acquisition, then you would have to propose a mechanism with which a learner can create language-specific connection from incoming sounds to phonological categories).

Category creation and constraint emergence were common features of Parallel Distributed Processing (Rumelhart, McClelland and the PDP Research Group 1986), but its child Optimality Theory did away with the distributedness and therefore lost the ability of creating categories and constraints. The same is true of that other child, Harmonic Grammar, which has a neurally inspired implementation (Soderstrom, Mathis & Smolensky 2006) that is not distributed either and therefore has to work with given categories and constraints (which are implemented as innate neural circuits). The only proposal for category creation in Optimality Thoery (Boersma, Escudero & Hayes 2003) had to work with costly computations of auditory distance, a feature typically found in exemplar models but not in connectionist models.

The present paper is a first attempt to bring distributed representations back into Optimality Theory. The result is a connectionist-inspired model (as OT is in general), without computations of auditory distance, which still has the capability of category emergence.

Apoussidou, Diana, & Paul Boersma (2004). Comparing two Optimality-Theoretic learning algorithms for Latin stress. *WCCFL* **23**: 29-42.

Apoussidou, Diana (2007). *The learnability of metrical phonology.* PhD thesis, University of Amsterdam.

Boersma, Paul (2003). Review of Tesar & Smolensky (2000): *Learnability in Optimality Theory*. *Phonology* **20**: 436-446.

Boersma, Paul (2008). Emergent ranking of faithfulness explains markedness and licensing by cue. *ROA* **954**.

Boersma, Paul (to appear). Some correct error-driven versions of the Constraint Demotion algorithm. To appear in *Linguistic Inquiry*. [*ROA* **980**]

Boersma, Paul, Paola Escudero, & Rachel Hayes (2003). Learning abstract phonological from auditory phonetic categories: An integrated model for the acquisition of language-specific sound categories. *Proceedings of the 15th International Congress of Phonetic Sciences*. 1013-1016. [*ROA* **585**]

Boersma, Paul & Bruce Hayes (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* **32**: 45-86.

Soderstrom, Melanie, Donald Mathis, & Paul Smolensky (2006). Abstract genomic encoding of Universal Grammar in Optimality Theory. In Paul Smolensky & Géraldine Legendre (eds.) *The harmonic mind*, Vol. 2. MIT Press. 403-471.

Rumelhart, David E., James L. McClelland, & the PDP Research Group (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. MIT Press.

Tesar, Bruce (1995). *Computational Optimality Theory*. Doctoral thesis, University of Colorado. [*ROA* **90**]

Tesar, Bruce (1997). An iterative strategy for learning metrical stress in Optimality Theory. In Elizabeth Hughes, Mary Hughes, & Annabel Greenhill (eds.), *Proceedings of the*

21st Annual Boston University Conference on Language Development, 615-626. Somerville, Mass.: Cascadilla. [*ROA* **177**]
Tesar, Bruce & Paul Smolensky (1993). The learnability of Optimality Theory: an algorithm and some basic complexity results. Ms. Department of Computer Science & Institute of Cognitive Science, University of Colorado at Boulder. [*ROA* **2**]
Tesar, Bruce, & Paul Smolensky (1998). Learnability in Optimality Theory. *Linguistic Inquiry* **29**: 229-268.
Tesar, Bruce, & Paul Smolensky (2000). *Learnability in Optimality Theory*. Cambridge, Mass.: MIT Press.

Çagri Çöltekin
c.coltekin@rug.nl
University of Groningen

Unsupervised Learning of a Highly Lexicalized Grammar

Mittwoch, 04.03.2009, 17.30 – 18.00

We present a model of acquisition of grammar. The model is based on a highly lexicalized theory of grammar, Combinatory Categorial Grammar (CCG). Two features of the CCG makes it an attractive formalism for the model of language acquisition presented here. First, the grammar of a language is completely specified in the lexicon, only a small set of universal rules that determine how adjacent categories merge and form higher level categories are left outside the lexicon. Second, CCG assumes close relationship between syntax and semantics. The former property helps combining lexical acquisition with grammar acquisition, which is also motivated psychologically (Bates & Goodman, 1997). The second property enables a learning system to make use of semantic information available in the input to learn grammar.

The use of (somewhat noisy and partial) semantic information is useful for a grammar learner and, arguably, this sort of semantic information is available to the children acquiring language. While learning from semantically annotated input in a computational learning system is one way to overcome discouraging results from learnability theory (Gold, 1967), this work is mainly focused on computational modeling of unsupervised learning of grammatical structures. More recent results from learning theory suggest that with reasonable constraints on the target grammar, formalisms based on Categorial Grammars are learnable using Gold's framework of learnability (e.g. Kanazawa, 1994).

The input to the model presented is unlabeled language data. The model generates all possible lexical hypotheses for each unit of input, and updates the lexicalized grammar incrementally using a simple Bayesian learning algorithm. The priors are based on the current lexicon, or uninformative for the unseen lexical hypotheses. The likelihoods favor short syntactic categories and a compact lexicon. We outline the theoretical foundations of the learning framework, followed by a discussion of learnability issues. We show that the system is capable of learning various grammatical phenomena. We also present results from the simulations where the model is applied to unsupervised learning of morphotactics of Turkish, a highly agglutinating language, using the data from child directed speech in the CHILDES database.

References

Bates, E. & Goodman, J. C. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. *Language and Cognitive Processes*, *15*(5/6), 507–584.

Gold, E. M. (1967). Language identification in the limit. *Information and Control*, *10*(5), 447–474.

Kanazawa, M. (1994). *Learnable Classes of Categorial Grammars*. PhD thesis, Stanford University.

Jos de Bruin*, Jacqueline van Kampen & Remko Scha*
jos.de.bruin@uva.nl, jacqueline.vankampen@let.uu.nl, scha@uva.nl
*ILLC, University of Amsterdam / UiL OTS, Utrecht University

Two acquisition paradoxes: A computational simulation

Donnerstag, 05.03.2009, 12.00 – 12.30

This paper presents initial results from a computational simulation of the development of child grammar. Grammar acquisition is modeled as a sequence of increasingly more complex grammars. A grammar is used both to parse the input language and as an evidence frame for subsequent extensions to that grammar. The grammars allow the child to reduce the input, focusing on a specific acquisition step, thereby avoiding the “Poverty of the Stimulus” problem. This means that the order in which various (adult language) constructions are acquired is determined not just by the frequencies with which these are used in the child's input, but also by the (child language) constructions the child recognizes in the input. This view on acquisition as a discovery procedure, rather than an evaluation procedure, has been used to explain unexpected acquisition orders in Evers & van Kampen (2001, 2008), Van Kampen (2004, 2006, 2007, 2008). Here we mention just two:

I. Germanic languages have parallel constructions for non-subject root questions.

(1) a. which bird do you see? b. welchen Vogel siehst du? c. welk vogeltje zie je?

These constructions begin with a wh-phrase followed by an inversion of finite verb and subject. English, a "residual V2" language, differs from other Germanic languages by allowing subject-verb inversion for only a small group of functional defective verbs. The other Germanic languages ("regular V2") allow inversion for any finite verb and moreover allow it for questions as well as topicalizations. In both systems, the subject-verb inversion indicates that the subject as an obligatory and constructionally marked argument has been captured and distinguished from the primitive notion “topic” (cf. Krifka 2006). One would expect that the primary learners of non-English are better prepared than the learners of English to acquire wh-words and inversion. CHILDES data (Miller, Groningen, Goteborg, Brown corpora) show that this is not the case. German, Dutch and Swedish children start to use V2 and inversion early, but delay the introduction of wh-words (Tracy 1994, Van Kampen 1997, Santelmann 1995). English children introduce wh-words early and drop the (residual) V2. Different primary systems (V2 Germanic, residual-V2 English) apparently invite different data-selections.

II. More than 2/3 (>75%, our count) of the non-questions in a V2 language start with an “aboutness” topic followed by a finite verb. Again, one would expect that the primary learners of a V2 language are well-prepared to acquire this “V2” property early. However, Dutch CHILDES data show that most (78%) of their early finite sentences have the finite verb not in V2, but in sentence-initial V1 position.

These two examples are explained by assuming a grammar-driven data reduction in early child language. This is confirmed by a computational simulation, which was tested on CHILDES input data. The model represents its evolving grammar as a collection of tree fragments and uses a generative mechanism for combining and transforming these trees. It thus combines techniques from Data-Oriented Parsing (a "usage-based model" which is formally a Stochastic Tree Substitution Grammar; Bod et al. 2003) with notions from classical Chomskyan syntax.

Roeland Hancock and Thomas G. Bever
{rhancock, tgb}@u.arizona.edu
Department of Psychology, University of Arizona, Tucson, AZ 85721 US

Extrinsic and Intrinsic Constraints on Language Acquisition – Language Acquisition as a Series of Revolutions

Mittwoch, 04.03.2009, 17.00 – 17.30

Two enduring problems have set limits on possible models of language acquisition: the poverty of the stimulus (POS); the creation of a new candidate grammar when an existing one fails (NCG). The hypothesis of innate "syntactic parameters" offers a common answer to both problems: the language acquisition device is pre-sensitized to data that determine each of a set of universal structural options. In its starkest form, this is a template trigger recognition theory of acquisition, not a learning theory.

Many mysteries emerge in how this abstract model works in an individual child: among others, recognition of each trigger presupposes prior structural analyses that put it into just the right parameter-changing context; – it is unclear how the individual child creates just the right ordering of parameter settings from fragmentary, unordered, often ungrammatical input that differs chaotically from child to child.

In light of such difficulties, we propose a dynamical systems model of acquisition in which the effects of parameter setting are obtained through statistical learning. On this model, the acquisition device discovers and stores representations of statistically preponderant meaning-form pairs in its experience – these serve as critical 'anchor points' for a dynamically converging algorithm that links settings on various parameters simultaneously, such that a change in one parameter setting immediately reconfigures settings on others into a new hypothesized grammar.

The variation in input from child to child leads to small and large changes in the behavior of the dynamical system. At the small scale, this yields idiosyncratic languages, characterized abstractly as a system functioning in the same qualitative regime as those of other speakers of the same typological language. At certain critical points, possibly linked to violations of the preponderant form, a shift in the statistical representations results in a radical change throughout the candidate grammar.

Three kinds of data can support this model. First, the model constrains properties of the actual language that the child experiences: it must offer statistically preponderant forms, both at the surface level and in the form->meaning mapping. Interestingly, attested languages exhibit this property: each language has a standard surface form and a typical mapping of that surface form onto thematic relations. On our view, many potential languages that conform to universal structural constraints are nonetheless unlearnable without such statistical eccentricities of form and usage.

Second, we will present a model of the historically attested shift from OV to VO. Unlike a reanalysis approach, our model can account for the initial rise in VO frequency and the relative lack of VO to OV shifts. The shift to VO is explained as a combination of a wider basin of attraction within the language system (without invoking Kayne's thesis of VO as a base order) and more restrictive OV anchor points in the statistical representations developed by a learner.

Finally, we have taken this model into the laboratory, studying its implications for learning a relatively naturalistic artificial language. We will report on specific results which support our model of language acquisition as occurring in revolutionary spurts, occasioned by successive building and then violation of statistical generalizations.

Gaja Jarosz
gaja.jarosz@yale.edu
Yale University

Effects of lexical frequency and phonotactics on learning of morphophonological alternations

Freitag, 06.03.2009, 12.00 – 12.30

Theories of learnability posit an early stage of phonotactic learning that aids in subsequent learning of morpho-phonological alternations (Prince & Tesar 2004; Hayes 2004; Jarosz 2006). Recent experimental work, however, indicates that even 4 yr old Dutch-learning children do not yet have productive knowledge of the morpho-phonological voicing alternation (Zamuner et al, in prep, "ZKF"; Zamuner et al, 2006), casting doubt on the role of phonotactics in subsequent learning. This paper demonstrates via simulation that the poor performance on alternating segments reported by ZKF is in fact expected in a computational model in which initial phonotactic learning precedes and aids morphophonological learning.

The phonotactic restriction on word-final voicing in Dutch leads to alternations for stem-final, voiced obstruents. ZKF show that the proportion of correct voicing in intervocalic coronal stops varies depending on underlying voicing, age, and whether the consonant alternates or is part of a monomorphemic word (Table 1A). Crucially, non-alternating /d/s are produced more accurately than alternating /d/s. If children employ phonotactic knowledge at this stage, what explains this difference?

The learning model, cast in a probabilistic Optimality-Theoretic framework, assumes an initial phonotactic-learning stage followed by a morpho-phonological learning stage during which underlying representations are gradually set and the grammar refined (Jarosz 2006). Since the model's task is to learn underlying forms as well as a constraint ranking, the data provided to the model is the distribution of stops (Table 2; data from ZKF) and their corresponding surface forms only. The model's predicted proportion of correct voicing is shown at two time steps (Table 1B), selected for their correspondence to the experimental age groups. There is close match between the predicted and experimental data, and crucially, non-alternating /d/s are produced more accurately than alternating /d/s by the model. Under this learning theory, the explanation for the observed effects lies in the frequency distribution of intervocalic stops. Because 95% of intervocalic stops in the input data are voiceless, the model predicts a strong initial grammatical preference for [t]. Further, because of the miniscule proportion of input data providing evidence of underlying voicing for alternating /d/s (0.5%), the model takes much longer to settle on /d/ for alternating /d/s than for non-alternating /d/s, which consistently surface as [d] (4.4%). Despite the availability of phonotactic knowledge, it is this *lexical* delay that results in low accuracy for alternating /d/s.

Low accuracy (in both the model and experimental data) of bi-morphemic /t/s relative to mono-morphemic /t/s provides the most direct evidence of the role of phonotactics. Both /t/s surface consistently as [t]s about 20% of the time. However, only bi-morphemic /t/s occur word-finally, and *if knowledge of word-final neutralization is employed,* word-final context provides no evidence of underlying voicing, leading to slower learning. In short, reliance on phonotactic knowledge can explain the lexical delay of bi-morphemic /t/s.

Table 1 - Proportion of Correct Intervocalic Voicing

A - Experimental Results from ZKF					*B - Computer Simulation*				
	bi-morphemic		monomorphemic			bi-morphemic		monomorphemic	
	/t/	/d/	/t/	/d/		/t/	/d/	/t/	/d/
Age 2;6	83%	5%	100%	28%	Iter. 112	87%	15%	90%	23%
Age 3;6	97%	13%	99%	75%	Iter. 126	96%	29%	99%	51%

Table 2 – Total Token Frequency of Stops by Type (from ZKF)

/t/→[t] (sg.)	*/t/→[t] (pl.)*	*/d/→[t] (sg.)*	*/d/→[d] (pl.)*	*monomorph. /t/*	*monomorph. /d/*
120 (11.8%)	87 (8.6%)	564 (55.5%)	5 (0.5%)	196 (19.3%)	45 (4.4%)

Janny Leung
(jannyleung@hku.hk),
The University of Hong Kong

Constraints in the Implicit Learnability of Form-Meaning Connections

Donnerstag, 05.03.2009, 12.30 – 13.00

I present in this paper empirical data from a set of experiments that used derived attention paradigms to examine implicit learning of associations between forms and meanings in a semi-artificial language. Studies of implicit learning have generally focused on the learning of rules and associations concerning forms in a single modality, for example letters in artificial grammar experiments, or spatial locations of visual stimuli in serial reaction time experiments. In the case of natural language systems learning such form-form associations might contribute to the acquisition of syntactic rules, but there remains the problem of learning how those forms relate to meanings. In my previous research, I have shown that people may become sensitive to associations between forms and meanings without an intention to learn them, and without awareness of what they are. However, they may only develop such sensitivity under certain conditions, such as when the meaning features are relevant to grammatical systems.

The experiments to be presented here examined whether some form-meaning connections were more easily learned implicitly than others, and whether this was attributable to the interactions of the form-meaning connections with grammar. The learning targets were the association of a set of determiners with animacy (Experiment 1A) and with relative size (Experiment 1B) respectively. The time taken to locate objects on the screen referred to by a noun phrase was a dependent measure. If subjects had learned the correlation between the determiners and the respective meaning, they might orient their attention to the correct referent once they heard the article and before they heard the noun. Learning was measured by slow-downs in search times when the learned associations were violated. Animacy is a grammaticalised concept but relative size is not. If implicit learning of form-meaning connections is confined to grammatised concepts then connections between functors and animacy should be implicitly learned, but connections between functors and relative size should not.

Selected bibliography:

Chun, M.M. & Jiang, Y. (1998) Contextual cuing: implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, 36, 28-71.

Doughty, C. & Williams, J. (1998) Pedagogical choices in focus on form. In C. Doughty & J. Williams (Eds.) *Focus on Form in Classroom Second Language Acquisition* (pp.197-261). Cambridge: Cambridge University Press.

Ellis, N. (1994) Vocabulary acquisition: The explicit ins and outs of explicit cognitive mediation. In N. C. Ellis (Ed.), *Implicit and Explicit Learning of Languages* (pp.211-282). London: Academic Press.

Gabrieli, J. D. E., Cohen, N. J., & Corkin, S. (1988) The impaired learning of semantic knowledge following medial temporal lobe resection. *Brain and Cognition*, 7, 157-177.

Gupta, P. & Dell, G. S. (1999) The emergence of language from serial order and procedural memory. In B. MacWhinney (Ed.), *The Emergence of Language* (pp.447-481). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Jiminez, L. (2002) Introduction: Attention to implicit learning. In L. Jiminez (Ed.), *Attention and Implicit Learning* (pp. 1-7). John Benjamins: Amsterdam.

Lambert, A. J. & Roser, M. (2001). Effects of bilateral color cues on visual orienting: revisiting William James' 'Derived Attention'. *New Zealand Journal of Psychology*, 30, 15-22.

Giorgio Magri,
gmagri@mit.edu
Dept. of Linguistics and Philosophy, MIT

A convergent version of the GLA which performs both promotion and demotion

Freitag, 06.03.2009, 11.30 – 12.00

The ranking problem (RP) in Optimality Theory (OT) is the problem of finding a hierarchy OT-compatible with a given comparative tableau (CT). Boersma's (1997) Gradual Learning Algorithm (GLA) for the RP iteratively updates a ranking vector by both promoting and demoting its components by a small fixed amount. The algorithm does not work, as shown by a counterexample recently found by Pater (2007). The problem is that the GLA performs promotion too, besides demotion. In fact, the variant with only demotion does work: it is a slowed-down version of Tesar's (1995) Constraint Demotion. Yet, we really want an algorithm that performs promotion too in order to model the early stage of the acquisition of phonotactics, as described in Hayes (2004). At this stage, the learner has no access to alternations and assumes that all underlying forms are fully faithful. Thus, the columns corresponding to faithfulness constraints in the CT contain no L's. Hence, a demotion-only algorithm never changes their initial ranking and thus cannot model the acquisition of phonotactics in this early stage.

I will offer an explanation of what exactly goes wrong when the GLA is run on Pater's counterexample. I will point out that the GLA fails on Pater's counterexample also when it is run deterministically with plasticity set to an arbitrary integer and a fractional random initial vector. Furthermore, I will provide a very simple counterexample that shows that the GLA fails also when run starting from the null initial vector. These considerations motivate the search for an alternative to the GLA.

My paper provides a provably convergent variant of the GLA that performs both promotion and demotion. The core idea is as follows. Without loss of generality, we can assume that every row of a given comparative tableau contains a single entry equal to L. Rewrite the entries W, L and E of CTs as 1, -1 and 0, respectively. We say that a CT **T** is *linearly compatible* (L-compatible) iff there exists a vector $\mathbf{w} = (w_1, \ldots, w_n)$ (where n is the number of columns of **T**) such that $w_1 t_1 + \ldots w_n t_n > 0$ for every row $\mathbf{t} = [\, t_1, \ldots, t_n \,]$ of the tableau **T**. It is known that OT-compatibility entails L-compatibility; and that the reverse is false. Yet, given a CT **T**, consider the tableau **T**' obtained from **T** by replacing any entry -1 with -k, where k is the number of entries +1 in the row where that negative entry sits. The following fact holds: **T** is OT-compatible iff **T**' is L-compatible; furthermore, if **w** is L-compatible with **T**', then every hierarchy which is a refinement of the possibly non-total hierarchy defined by the ordering of the components of **w** is OT-compatible with **T**.
This claim says that the RP in standard OT reduces to a linear feasibility problem and that we can thus import within OT methods and results from the rich literature on linear classifiers. In conclusion, we straightforwardly get a convergent version of the GLA by running the Perceptron (or more modern variants thereof) on the table **T**'. I will consider two such algorithms. These algorithms perform both promotion and demotion, as desired. I will thus show that they are suitable to model the acquisition of phonotactics in the early stage described by Hayes (2004).

Frédéric Mailhot
fmailhot@connect.carleton.ca
Carleton University

Inducing vowel harmony from iterated learning over noisy data with Bayesian agents

Mittwoch, 04.03.2009, 15.30 – 16.00

It is well-known that vocalic properties are coarticulated across an intervening consonant (*viz.* in VCV sequences) in running speech (Hardcastle & Hewlett 1999). We present ongoing work investigating Ohala's (1994) claim that vowel harmony develops as the phonologization of this vowel-to-vowel coarticulation, *i.e.* it is a form of sound change.

We test this theory in the context of an agent-based model using the *iterated learning* paradigm of Kirby (1998), in which a population of agents is divided into “adults” and “children” with the learned phonological forms of each generation serving as the inputs to learning for the successive one. Adults output noisy acoustic forms of a simplified lexicon (only [±High] and [±Back] vowels are modeled) using a realistic articulatory synthesizer, which is coupled with articulatory undershoot and variation along the effective second formant (F2'), which is taken to be a suitable analogue of front/back coarticulation.

We adopt an “ideal observer” stance, modeling our children as optimal Bayesian learners whose task is to assign structural descriptions (phonological forms) to the acoustically-encoded data. The hypothesis space over which inference is performed includes 2 binary phonological features, and we examine the effects of uniform priors as well as priors structured by the physiological constraints on production.

The overarching goal then, is to examine by means of a specific case, how the dynamics of iterated learning of phonological forms evolve under the assumption of optimal inference from noisy data.

References

[1] Hardcastle, W.J. and N. Hewlett (1999). *Coarticulation: Theory, Data and Techniques.* Cambridge University Press.
[2] Kirby, S. (1998). *Function, Selection and Innateness.* Oxford University Press.
[3] Ohala, J. (1994). Towards a universal, phonetically-based, theory of vowel harmony. ICSLP 3, Yokohama, pp. 491-494.

Stefan Müller
Stefan.Mueller@fu-berlin.de
Freie Universität Berlin

Learnability of Phrasal Patterns

Donnerstag, 05.03.2009, 09.30 – 10.00

Linguistic theories differ in their assumptions regarding the nature of grammatical constructions. LFG, HPSG, Categorial Grammar, and LTAG assume that valency information is encoded in the lexicon and is projected from there. They assume that active/passive alternations are treated by lexical rules that relate stems to active and passive lexical items or–in the case of LTAG–elementary trees. As an example consider the resultative construction in (1b):

(1) a. Peter fishes. b. Peter fishes the pond empty.

Proponents of lexical theories assume a lexical item that differs from the one used for (1a) by selceting an additional object and a predicate. In contrast most proponents of Construction Grammar assume phrasal constructions, that is, they assume a pattern for the resultative construction into which the verb fish can enter. The resultative meaning only arises if the respective lexical items are realized in the resultative construction. Instead of assuming two lexical items and very general combination schemata, Construction Grammarians assume special combination schemata and general lexical items (Goldberg, 1995; Goldberg and Jackendoff, 2004). As Müller (2006) points out, the phrasal approaches have problems accounting for the interaction with other parts of the grammar. For instance active/passive alternations and similar valence changing processes cannot be accounted for in purely phrasal treatments that do neither assume transformations nor lexical rules. Valence changing processes like causation that can be iterated cannot be modeled in inheritance hierarchies.

In my talk I will discuss the relation of phrasal and lexical approaches in terms of learnability. I show that language acquisition arguments for phrasal constructions are not convincing, since the facts can be explained in lexical models as well. For instance a phrasal construction of the form [Subj [V Obj]] can be modeled by a Categorial Grammar lexical entry of the form (s\np)/np. The lexical treatment has the advantage that it is easy to explain why adjuncts can intervene between the subject and the VP.

I will discuss the phrasal, inheritance-based approach to Persian Complex Predicates that was suggested by Goldberg (2003) and show that it requires that each interaction between a complex predicate and other constructions (negation, clitics, future) would have to be acquired independently if her approach would be correct. As Tomasello (2000) has shown, this is not realistic in terms of language acquisition, since competent speakers of a language can use acquired rules for newly acquired verbs.

Goldberg, A. E. 1995. Constructions. A Construction Grammar Approach to Argument Structure. Chicago/London: University of Chicago Press.
Goldberg, A. E. 2003. Words by Default: the Persian Complex Predicate Construction. In E. J. Francis and L. A. Michaelis (eds.), Mismatch: Form-function Incongruity and the Architecture of Grammar, pp 117–146, Stanford: CSLI Publications.
Goldberg, A. E. and Jackendoff, R. S. 2004. The English Resultative as a Family of Constructions. Language 80(3), 532–568.
Müller, S. 2006. Phrasal or Lexical Constructions? Language 82(4), 850–883.
Tomasello, M. 2000. Do Young Children Have Adult Syntactic Competence? Cognition 74, 209–253.

Maren Pannemann
mpannemann@hotmail.com
Radboud University Nijmegen

Determiner Phrase Acquisition at the Syntax-Phonology Interface

Donnerstag, 05.03.2009, 10.00 – 10.30

A characteristic difference between adult and child language is that children omit functional elements such as determiners. Instead, they pass through a stage with lexical words only, where they produce bare nouns such as “ball!” or “doggy!”. From the mid 90’s onwards, several generative studies have aimed to account for the absence of determiners in terms of a bottom-up structure building (SB) model of syntax acquisition: initially, only the lexical projection N is represented in the children’s grammar (cf. Clahsen et al 1994, among others). Only in a subsequent stage, the phrase structure expands with a functional projection hosting the determiner. The aim of the present talk is twofold: First, I will argue that the SB model is challenged by learnability considerations and illustrate this with first language(L1) acquisition data. Second, I will present an alternative approach to DP acquisition in line with learnability.

Within the SB model, the child has no other option than to assume that all nominal items correspond to the lexical category N at the onset of acquisition. This hypothesis proves to be correct for common nouns like “house”. However, it is not correct for proper names, like “Mary”, and pronouns like “he”, which have a different syntactic status as -in languages like English, French or Dutch- they cannot be modified by a determiner. According to the *subset principle* (Berwick 1985, Wexler and Manzini 1987), the child can give up a hypothesis only in the presence of positive evidence. However, pronouns and proper names never combine with a determiner in the child’s input. Hence the child would have to conclude from the mere absence of a construction that it is not possible, a view that conflicts with the subset principle. Furthermore, if indeed children initially hypothesised that proper names and pronouns are Ns, we would expect that children accidentally combine them with determiners. However, as I will show, these kinds of errors cannot be observed. A longitudinal study of spontaneous speech data in two French and two Dutch children (age range 1;07–3;3, with a total of 886 instances of proper names and 9090 pronouns) reveals that -unexpected in the SB model- young children do not combine proper names and pronouns with determiners.

In the present talk I will present an alternative approach to DP acquisition that can account for this observation and is in line with the subset principle. With Halle and Marantz (1993) and others I assume that phonological material is associated with syntactic representations at the PF interface, and that *correspondence rules* (Jackendoff 1997) connect phonological units with their syntactic representation. One consequence is that phonological words can correspond to maximal projections. In fact, the subset principle dictates that at the onset of acquisition all nominal items correspond to the maximal projection DP. The child has to reset this initial hypothesis for common nouns, and positive evidence in the input allows her to do so, as common nouns do overtly combine with determiners. In this sense, acquisition of lexical words and their syntactic categories goes hand in hand and can be characterised as a process of resetting correspondence rules at the PF interface level. Driven by positive evidence only, DP structure is acquired in a top-down process. As the functional D domain of DP is available from the onset of acquisition, common nouns (like “ball!”) can refer to specific entities even in the absence of an overt determiner. As I will discuss in the remainder of the talk, this setting is compatible with the psycho-developmental reality of L1 acquisition (*Theory of Mind*).

Lisa Pearl
(lpearl@uci.edu)
University of California, Irvine

Parametric Linguistic Systems: The Limits of Probabilistic Learning for Realistic Data

Donnerstag, 05.03.2009, 09.00 – 09.30

A strength of probabilistic learning is that it often allows children to make generalizations from ambiguous and noisy data effectively. Here, we focus on how young children could acquire a parametric system of metrical phonology (adapted from Dresher (1999) and Hayes (1995)), using English as a case study. We examine several cognitively plausible probabilistic learning models and find that, given empirical data as input, purely probabilistic models fail to converge on the English grammar reliably while models with a selective learning bias succeed.

The learning task for English metrical phonology is not trivial; the system explored here has 9 interacting parameters, making it difficult to determine which grammar generated a given data point. Moreover, the English data are highly noisy. Data from the Bernstein and Brent corpora in the CHILDES database (MacWhinney 2000) suggest that over 27% of the data are incompatible with the English grammar for at least one parameter value. Together, these factors make the learnability of the English metrical phonology system an interesting case study.

The unbiased probabilistic learning models examined here are adapted from the cognitively plausible model in Yang (2002), and assume children process data incrementally. Several variations of incremental learning models are tried, including those with batch learning and prior knowledge of part of the English system. Yet all these models fail. Close examination of the data reveals something that will trouble *any* unbiased probabilistic learner: the correct English grammar is not the one most compatible with the English child-directed speech data. So, a purely probabilistic learner would be unlikely to choose it.

Still, since English children seem to learn the English grammar, there must be some bias they bring to the learning scenario that makes the English grammar the most compatible. One bias that does lead to the correct learning behavior is to learn only from the subset of the available input that is perceived as *unambiguous* (Pearl 2008). When a learner uses this selective learning bias, the English grammar is learned from the English data.

While this may not be the only successful bias, the results here suggest an unbiased probabilistic learner will flounder when given realistic English data. Previous results for English provide one learning bias that will generate the correct learning behavior. Because the correct grammar *can* be learned -by biased learners- from the empirical data children use as input, this supports the parameter system as a viable hypothesis space instantiation that children could use.

Dresher, E. (1999). Charting the learning path: Cues to parameter setting. *LI 30*, 27-67.
Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.
MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. Mahwah, NJ: Lawrence Erlbaum Associates.
Pearl, L. (2008). Putting the Emphasis on Unambiguous: The Feasibility of Data Filtering for Learning English Metrical Phonology, in Chan, H., Jacob, H., and Kapia, E (eds.), *BUCLD 32.* Somerville, MA: Cascadilla Press, 390-401.
Yang, C. (2002). *Knowledge and Learning in Natural Language.* Oxford: Oxford University Press.

Bruce Tesar
tesar@rutgers.edu
Department of Linguistics, Rutgers University, New Brunswick

Learning Phonological Grammars for Output-Driven Maps

Mittwoch, 04.03.2009, 14.00 – 15.00

The challenge of simultaneously learning a lexicon of underlying forms and a constraint ranking has been addressed by several scholars in recent work (Apoussidou 2007, Jarosz 2006, Merchant 2008, Tesar 2006). While making significant progress, each of these proposals includes computational demands that are unlikely to scale up in a cognitively plausible fashion.

Merchant's proposal, the Contrast Pair and Ranking information algorithm (CPR), faces the following computational difficulties. (1) The focus of CPR on lexical hypotheses for only a pair of related words at a time (a contrast pair) is a vast improvement over simultaneous consideration of all possible lexica, but the space of lexical hypotheses for a single contrast pair still grows exponentially in the number of unset underlying features for the morphemes involved in the pair. (2) The extraction of ranking information performed by CPR is able to obtain ranking information from contrast pairs for whichcomplete underlying forms have not yet been determined, but faces exponential computationalcomplexity. (3) The technique of initial lexicon construction, setting in advance features that do notalternate across the entire paradigm, can restrict further the number of lexical hypotheses that need to be considered, but at the cost of requiring that the learner have a complete paradigm of surface form data before learning of underlying forms can begin.

This talk will demonstrate that each of these computational concerns can be significantly improved upon by taking additional grammatical structure into consideration. The key additional structure lies in the concept of an output-driven map (Tesar 2008). Intuitively, an output-driven map is a phonological map in which all disparities introduced between the input and the output are motivated by conditions on the output. While the intuition is a familiar issue in theoretical phonology, the concept of output-driven map formalizes it in a novel way by requiring that any grammatical input-output mapping A→C entails the grammaticality of B→C whenever B has greater similarity to C than A does (A→C has every input-output disparity that B→C does, but B→C may lack some disparities of A→C).

The structure of output-driven maps can be exploited in learning via the contrapositive: the ungrammaticality of B→C entails the ungrammaticality of A→C. Given a grammatical output C, it is necessarily the case in an output-driven map that C→C (the restricted identity map property). Suppose B→C has one disparity (e.g., they differ in the value of one feature on one segment). If the learner possesses sufficient information about the grammar to determine that B→C is ungrammatical, then the learner need not bother checking any A→C that includes the disparity of B→C; because the map is output-driven, the ungrammaticality of A→C is entailed by the ungrammaticality of B→C. Instead of needing to evaluate all combinations of possible values for all unset features of a word (exponential in the number of unset features), the learner can obtain the same information while only evaluating a single unset feature at a time (linear in the number of unset features), having the other unset features match (temporarily) the values of their output correspondents, addressing concern (1). Even greater benefit is realized when obtaining ranking information from forms with unset features, addressing concern (2).

The speed-up realized by exploiting the structure of output-driven maps is significant enough that initial lexicon construction is no longer needed. This frees the learner from needing an entire paradigm before learning commences; the learner can begin learning about underlying forms from even a single datum, addressing concern (3). This algorithm has the notable property that features of underlying forms which cannot be shown to require a particular value remain unset; non-contrastive features are never set, without any need for the learner to separately construct an "inventory of contrastive features".

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.

Tesar, Bruce. 2006. Learning from paradigmatic information. In *Proceedings of the 36th Meeting of the North East Linguistics Society*, ed. by Christopher Davis, Amy Rose Deal, and Youri Zabbal, 619-638. GLSA. ROA-795.

Tesar, Bruce. 2008. Output-driven maps. Ms., Linguistics Dept., Rutgers University. ROA-956.

Jacolien van Rij-Tange[1,2], Petra Hendriks[1], Hedderik van Rijn[2]
[1] *Center for Language and Cognition Groningen, University of Groningen*
[2] *Department of Artificial Intelligence, University of Groningen*
Email addresses: J.C.van.Rij@rug.nl, P.Hendriks@rug.nl, D.H.van.Rijn@rug.nl

Using cognitive models to study the acquisition of pronoun comprehension

Donnerstag, 05.03.2009, 11.30 – 12.00

In this study the acquisition of pronoun comprehension has been investigated with the use of computational models. Up to the age of 6, children have been shown to experience difficulties in the interpretation of pronouns (but not reflexives) by incorrectly allowing the pronoun to corefer with the local subject about half the time (e.g., Chien & Wexler, 1990). Hendriks, van Rijn & Valkenier (2007) argue that this Delay of Principle B-Effect (DPBE) arises because children are unable to take into account the speaker's perspective, operationalized as bidirectional optimization in the Optimality Theory framework. In their computational ACT-R (Anderson et al., 2004) model of the development of pronoun comprehension proficient pronoun comprehension is simulated as two consecutive steps of unidirectional optimization. However, if the speed of processing is not fast enough to perform both steps within a limited amount of time, pronouns remain ambiguous and a guessing pattern emerges. Assuming that children's performance is slower than adult performance, the ACT-R model predicts that children's performance will improve if they are given sufficient time to construct an interpretation, for example by decreasing the speech rate.

This prediction is tested by comparing the performance of 62 children (age 4;1-6;2, mean 5;1) on pronoun comprehension at a normal rate (4.0 syll/sec) with slow speech (2.7 syll/sec). It was found that slowed-down speech has a significant beneficial effect on children's pronoun comprehension, but only if the child displays a DPBE. Furthermore, the ACT-R model of learning bidirectional optimization is integrated in a computational model of sentence processing (Lewis & Vasishth, 2005) to simulate children's behavior on the experiment. The resulting ACT-R model processes sentences word by word, so that the effects of different presentation word rates could be modeled. The ACT-R simulation of the experiment shows a good fit of the data of children who show the DPBE.

The results of the experiment and the simulation support the hypothesis that children's difficulties with pronoun comprehension are caused by their limited speed of processing.

References.

Anderson, J., Bothell, D., Byrne, M., Douglass, S., Lebiere, C. & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111*(4), 1036-1060.

Chien, Y. & Wexler, K. (1990). Children's Knowledge of Locality Conditions in Binding as Evidence for the Modularity of Syntax and Pragmatics. *Language Acquisition, 1*(3), 225-295.

Hendriks, P., Van Rijn, H. & Valkenier, B. (2007). Learning to reason about speakers alternatives in sentence comprehension: A computational account. *Lingua, 117*(11), 1879-1896.

Lewis, R. & Vasishth, S. (2005). An Activation-Based Model of Sentence Processing as Skilled Memory Retrieval. *Cognitive Science, 29*(3), 375-419.

Aleksandra Zaba
aleksandra.zaba@uni-hamburg.de
University of Hamburg

On the relation between cognitive biases, learnability, and typological frequency: The case of vowel and consonant harmony

Donnerstag, 05.03.2009, 10.30 – 11.00

The cross-linguistic frequency of patterns is commonly linked to their learnability, in that learnability causes frequency based on cognitive biases (Chomsky & Halle 1968:4, 251, 296-297; Steriade 2001:235-237). Recently, researchers have used experimental methods to test this claim. Wilson (2003) found that the frequently attested pattern of nasal consonant harmony (example: nonce word /dume-l̲a/, which becomes [dumen̲a] vs. /tuko-l̲a/, which becomes [tukol̲a] since no nasal consonant is present in the stem), is more learnable by adults than are unattested patterns (cf. Koo & Cole 2006).

The present study investigated whether the relative frequency scale suggested in Hansson (2001), in which back vowel harmony (BH) (example: [Sim…Sim-æd] vs. [®Ud…®Ud-Ad]; stem /U/ turns suffix /æ/ to /A/) >> nasal consonant harmony (NH) (example: [®Ud…®Ud-æd] vs. [Sim…Sim-æn]; stem /m/ turns suffix /d/ to /n/) >> labial consonant harmony (LH) (example: [®Ud…®Ud-æd] vs. [Sim…Sim-æb]; stem /m/ turns suffix /d/ to /b/) corresponds to the same learnability scale. Participants with no knowledge of harmony languages were randomly assigned to the BH, NH, or LH training condition (n=19 per condition), auditorily exposed to words in an artificial language, and instructed to (implicitly) learn their respective word formation pattern. Example words in the BH condition include [fEn…fE̲n-æ̲d] and [sub…su̲b-A̲d], in NH [sub…sub̲-æd̲] and [fEn…fEn̲-æn̲], and in LH [fEn…fEn̲-æd̲] and [sub…sub̲-æb̲]. Following training, participants were tested on their knowledge of the patterns. The predicted learning scale was BH>>NH>>LH. Relevant test words contained stems that were unfamiliar from the training (‘new’ words). Some of these words followed the trained pattern (‘new grammatical’ words), others did not (‘new ungrammatical’ words).

To investigate learning, d’-scores were calculated, which allowed us to measure participants’ ability to distinguish between grammatical and ungrammatical words. The rationale was that, if the BH condition was better than the other two conditions at discriminating between new grammatical and new ungrammatical words, and the NH condition was better than the LH condition (i.e., if BH>>NH>>LH in their ability to extend their knowledge of the trained pattern to words with new stems), then the learning scale could be said to have been BH>>NH>>LH. Preliminary analyses showed no significant differences in d’-scores among the conditions. To exclude the possibility that these results were due to individual differences in learning ability, the analysis was repeated, but using only participants with a d’-score of 0 and above (see also Kates, Krauss, AbdulSabur, Colgan, Antshel, Higgins & Shprintzen 2007). Results revealed no significant differences in d’-scores among the conditions. Specifically, there were no significant differences in d’-scores between BH (.64) and LH (1.19) ($F(1,23)=1.502$, $p=.233$, partial eta squared=.064), nor between NH (1.02) and LH ($F(1,24)=.128$, $p=.724$, partial eta squared=.006), nor between BH and NH ($F(1,26)=2.921$, $p=.100$, partial eta squared=.105). Thus, no support was provided for the hypothesis that BH>>NH>>LH condition in learning. Possible explanations include that for some patterns, frequency is based on other factors than learnability and cognitive biases.