Précis of *Modeling words in the mind*

Constantine Lignos

Introduction

If the language we produced consisted solely of the repetition of sentences we had heard before, the study of language would indeed be a boring practice. But language is much more diverse than this; we are able to combine some units—building blocks, be they words, morphemes, or some other sort of linguistic atoms—to make "infinite employment of finite means" (von Humboldt, 1836/1988) to communicate, enabling the study of the cognitive processes behind this feat.

This dissertation uses computational modeling to explore these units through the study of language acquisition and processing, building computational models that improve our understanding of the representation and structure of words in the mind. I take an interdisciplinary approach to the development of these models, uniting perspectives from linguistic, cognitive, and computational inquiry to gain a holistic understanding of the cognitive processes involved. The models are tested against data from diverse sources, including corpora of child-directed speech (CHILDES; MacWhinney, 2000), a "megastudy"-scale lexical decision experiment (Balota et al., 2007), and evidence of historical changes in English pronunciation (Elphinston, 1765).

Specifically, this dissertation uses computational modeling to address three questions related to the units of language acquisition and processing:

1. How do children learn to divide the speech stream into appropriately-sized units?
2. How are these units represented and accessed in the adult mind?
3. How are generalizations about the regular and irregular characteristics of these units represented and learned?

Throughout this work, language acquisition and processing are treated as two views of the same cognitive mechanisms. As demonstrated through acquisition error patterns (Chapter 3), adult lexical decision times (Chapters 4 and 5), and historical changes in the pronunciations of words (Chapter 5), in both learning and processing the human mind takes a *decompositional* approach. It eagerly splits utterances into words, words into component morphemes, and surface forms into the most parsimonious phonological derivations possible. By building detailed computational models of these processes, I am able to show that a single memory retrieval mechanism, serial access (Murray and Forster, 2004), provides an explanation for many of the observed behaviors and a unifying link between the processing of regular and irregular linguistic forms.

In addition to modeling these specific cognitive processes, this dissertation seeks to advance the practice of building computational models of cognitive processes. While statistical methods, modeling techniques, and data availability have changed dramatically over the last thirty years, there has been little discussion of the methodology of relating computational models and cognitive processes since the development of the levels of analysis framework proposed by Marr (1982/2010). Before addressing the above questions, I revisit Marr's framework, discuss its applications to modeling language in the mind, and demonstrate the importance of attending carefully to levels of analysis when considering experimental results.

The studies in this thesis develop the simplest computational models possible from the ground up, eschewing complex frameworks in the quest for simple, interpretable models that fall out directly from the behavioral data. By advancing a computationally minimalist approach to cognitive modeling, I demonstrate that even very simple models can shed light on long-standing debates when paired with large-scale data. Simple models and big data can yield great dividends when carefully aligned with existing findings from experimental and theoretical linguistics and cognitive science.

Building computational models of cognitive processes

In our quest to understand cognitive processes, we have many empirical methods at our disposal. Why should we invest effort in building computational models? Chapter 2 has two goals: to establish a framework that will allow for discussion of the relationship between computational models and cognitive processes, and to demonstrate the importance of computational modeling in our study of cognition. I argue that the primary virtue of computational modeling is its ability to provide a testing ground for theories of cognition with a greater level of control than is available through observational or experimental approaches alone. When properly paired with behavioral evidence and accompanying theories, building and evaluating computational models allows us to explicitly test the proposed mechanisms of cognition.

I discuss what defines a computational model of cognition and how it may differ from an engineering solution to a similar problem. To establish a consistent terminology and approach to characterizing computational models, I review Marr's *levels of analysis* (Marr, 1982/2010) with a focus on the distinction between computational-level and algorithmic-level models. While a computational-level model defines the desired input/output property of a system and the goals it tries to optimize, an algorithmic-level model must provide a specific, implementable mechanism for achieving that goal. Due to the increasing popularity of statistical optimization techniques, many popular models in the domains explored in this thesis are defined only at the computational level, leaving the algorithmic level open for further study.

However, there are substantial differences between the relatively stateless information-processing task of visual processing that Marr explored and the stateful problems of language acquisition and processing explored in this thesis. I propose requirements for what must be specified for a learning model to be a valid model at the computational, algorithmic, and physical levels. This facilitates the discussion of models in the following chapters, where I propose algorithmic-level models of word segmentation and lexical processing.

After reviewing and extending Marr's levels of analysis framework, I demonstrate how computational modeling can be an essential companion to human subjects studies. To show the importance of building explicit computational models of learning processes and testing them, the remainder of this chapter examines the performance of participants in a language learning study. I use an artificial language syntactic learning experiment (Saffran, 2001) as a case study to demonstrate the value of using computational modeling even for simple psycholinguistic experiments.

I demonstrate that comparing participant performance against trivial computational models reveals that there is little support for the author's conclusions as stated; the author claims participants learned to identify hierarchical structure, but simulation shows that no knowledge of hierarchical structure is required to demonstrate the type of discrimination shown by human subjects. Furthermore, the error pattern of study participants is similar to that of one the non-hierarchical learning strategies modeled, suggesting that it provides a better explanation of the study results than assuming participants learned a hierarchical representation of the artificial language. This highlights the importance of building accompanying baseline learning models to evaluate participant performance.

I argue that participant performance in a learning experiment cannot be used to determine the details of *what* and *how* the participants learned without careful comparison of their performance to computational models; to do so would be drawing conclusions about the computational level without verifying that the description at that level is adequate. I name the issue of evaluating experimental results without an explicit computational model *the poverty of the experiment*, and argue that experiments that seek to provide evidence of a specific cognitive mechanism must model alternatives to show that there is indeed specific evidence for the mechanism proposed.

Word segmentation

With a firm basis in the theory of computational cognitive modeling established, the thesis now turns to the first of the problems being modeled. For adults, the decomposition of continuous speech into words—a problem known as word segmentation—is effortless. However, for an infant with limited knowledge of the language being learned, the task is far from trivial. To succeed, she must identify the words of her language from a stream of input that provides no pauses between words, few words in isolation, and little feedback.

Chapter 3 presents a model of infant word segmentation, developing an algorithmic-level model that mirrors the developmental trajectory of children as they learn to segment speech. I discuss previous experimental and computational findings in word segmentation and propose a model that addresses a previously unmodeled phenomenon: while learning to segment words, learners progress from treating oversized units as words—word collocations and sometimes phrases—to undersized ones, for example segmenting *behave* as *be* [heɪv] (Brown, 1973; Peters, 1983). The model proposed is computationally simpler than many previous approaches yet is able to segment accurately when evaluated against child-directed speech corpora. The model's primary mechanism is aggressive decomposition, breaking the input into as many pieces as appears to be useful, a mechanism which is revisited in Chapter 4, where similar behavior is found in online lexical processing.

Chapter 3 begins with a review of experimental and developmental studies of infant word segmentation. I identify the gap between the popular infant and adult experimental studies in this area (e.g., Saffran et al., 1996) and the performance of computational models derived from the mechanisms those studies propose (Yang, 2004). While experimental work has provided insight into the types of cues infants *might* be using, computational modeling of the task provides a unique opportunity to test proposed cues on representative data and validate potential approaches to using them. A faithful cognitive model of word segmentation must integrate knowledge of infants' experimental performance and developmental trajectory with a computational modeling framework to build as complete a picture of word segmentation as possible.

While there is clear benefit to aligning computational models of word segmentation with human subjects data, the connection between computational models and experimental and longitudinal studies of word segmentation is thus far tenuous. The most well-known modern computational models (e.g., Goldwater et al., 2009; Johnson and Goldwater, 2009) heavily focus on the application of ideal learner models but have not connected with infant development in a substantive way. In those models and many others, the focus has been on Marr's computational level of analysis, investigating *what* a learner may be trying to optimize when trying to learn to segment rather than *how* it may reach its goal.

In contrast, the approach I propose in this chapter focuses on building a model explicitly at Marr's *algorithmic* level, developing a specific, cognitively plausible strategy based on infants' performance in word segmentation experiments and comparing its predictions to the patterns observed in infant development. I use experimental evidence to suggest the operations the segmenter should perform and propose a simple learning algorithm that predicts the type of developmental changes observed in child language development. I demonstrate the effectiveness of this model at the task of word segmentation and at replicating patterns of infant development.

Processing of morphological structure

When we try to retrieve a word like *played* from memory, on what units are we performing lexical access? We may simply treat it as a single atom, accessing it as a *whole word* and assuming it is stored as such. We may treat it as a combination of *play* and *-ed*, dividing it under a model of *obligatory decomposition* into smaller pieces for lexical access. We may use a combination of both strategies, exercising multiple routes for lexical access, with each route optimized for certain types of words. This question has been discussed heavily from theoretical and experimental perspectives but has received little comprehensive treatment using computational methods.

Chapter 4 develops and compares word structure-aware models of lexical access from the ground up in the largest-scale computational study to date, testing the predictions models of lexical access on a large corpus of lexical decision times. As in Chapter 3, the study here begins by first trying to establish the computational-level parameters of lexical access of morphologically complex items and then explores algorithmic-level mechanisms that can give rise to these behaviors.

I first review the impasse that has formed around models of processing morphologically complex forms. I use *frequency effects*, the facilitation of access for more frequent units of mental representation, in a large corpus of lexical decision task trials (English Lexicon Project; Balota et al., 2007) as a diagnostic to identify how words are represented in the mind. Contrary to previous claims (Alegre and Gordon, 1999; Baayen et al., 2007), I find relatively uniform frequency effects across a broad range of frequencies, suggesting that there is no discontinuity between the representation of high and low-frequency items as proposed by some dual-route models. I find that assuming obligatory decomposition of words into morphemes regardless of word frequency—a single-route model—provides the best model of behavioral data.

To provide an algorithmic explanation of these frequency effects, I test the *rank hypothesis* (Murray and Forster, 2004), finding that the simplest explanation for these effects is simply that retrieval of lexical items proceeds as if traversing a long list, a *serial access* mechanism. I use this finding in combination with that of decomposition to support a model of lexical processing in which words are always decomposed into their component morphemes. In this model, access time is proportional to the rank of the word's base and the probability of the base combining with affixes contained in the word.

Using computational modeling of reaction time data, I find that the following simple decompositional model provides the best explanation of the data:

1. Morphologically complex words are decomposed into roots and affixes
2. The speed with which complex words are recognized is proportional to:
    (a) The rank of the root, due to serial access mechanisms
    (b) The probability of the root taking the specified affixes

In addition to explicitly testing the predictions of this model, I address other proposals, namely that for some particular—often underspecified—set of words, there will be no evidence of decomposition. I demonstrate that high frequency words participate no less in base frequency effects than low frequency ones (*contra* Alegre and Gordon, 1999), and words that are of higher frequency than their bare form have no special status that causes them to be stored and accessed as whole words (*contra* Hay and Baayen, 2002). I discuss the difficulty in assessing claims made by other dual-route proposals, such as those made by Baayen et al. (1997), which often give a general criterion but no reproducible way to test it given a new set of items under study. After assessing alternative dual-route models, I conclude that the best model for the processing of regularly inflected forms is the one proposed in this chapter: a decompositional, single-route approach built on serial access mechanisms.

The representation of generalizations over words

Chapter 5 explores the implications of the algorithmic-level model of lexical processing identified in Chapter 4 on the representation of irregular forms and on language change. Building on a formal productivity-based model of irregular representation (Yang, 2005) that depends on serial access assumptions, I explore how irregular forms are accessed and how that representation may change over time as the input is optimized for processing through generalization.

Using lexical decision times from the English Lexicon Project, I find that the same serial access mechanisms that provide the best model of the processing of regularly-inflected forms also explain the processing of irregulars when the productivity model of Yang 2005 is assumed. As predicted by the productivity model, I find evidence of serial access across classes of irregulars. This provides the first explicit test of the specific mechanisms of the productivity model on behavioral data and adds evidence for the psychological reality of classes of irregular verbs.

I propose that the generalization criterion suggested by the productivity model can be used as a test for what generalizations learners will acquire from the input by applying this test over simulated generations to simulate *analogical change*, historical change regarding the broadening of the scope of a linguistic generalization. I evaluate this model on the *postnasal stop deletion* change, a historical change in the English language in which the pronunciation of words like *king* changed from [kɪŋg] to [kɪŋ] through deletion of the final /g/. This change is known as a *domain-narrowing* change because over time it proceeded from first applying to only phrasal units, then to individual words, and finally to applying to stems inside of words.

To model an analogical change, I apply productivity as a test for what generalizations learners will acquire over simulated generations of learning. I propose several strategies for dealing with the variation across individual tokens that previous applications of productivity have not had to address, with these strategies varying in how aggressively they promote change. In the simulation of the postnasal stop deletion change, I find that both an aggressively change-promoting strategy and specific properties of the English language—restricted phrase-level resyllabification and frequent surfacing of bare stems—were necessary for the change to have proceeded in the observed historical pattern. The study of language change presented in this chapter represents several firsts: the first explicit model of a domain-narrowing change, the first application of productivity to inter-generational learning, and the first exploration of strategies for identifying forms as participants or exceptions in a change when the input show conflicting patterns.

Conclusions and interdisciplinary contributions

The most clear conclusion to be drawn from the effectiveness of the cognitive models in this dissertation that simple computational models can explain much of what there is to explain, often more than more complex and less interpretable models can. The research program developed throughout this dissertation focuses on simple, testable, and falsifiable computational models for the domains discussed, while trying to carefully justify the methodology under which they are explored. While there are many alternative models for the problems described in this dissertation, few other proposed models place themselves so carefully within both experimental findings and the broader context of the philosophy of cognitive modeling. I believe that it is this holistic view which separates the models presented in this dissertation from previous work and holds the most promise for future modeling.

A consistent theme throughout this dissertation is that there is great benefit to adopting a unified approach to the modeling of language acquisition and processing. This is in many ways a natural extension of progressing to modeling at the algorithmic level; once we begin to define cognitive processes using explicit procedures, we can start to see commonalities that may have been missed. For example, the heuristic choice in subtractive segmentation in the word segmentation model in Chapter 3 of "most frequent word wins" can straightforwardly be implemented by the serial access mechanism at the core of the lexical processing model in Chapters 4 and 5; the most frequent word will be the first one found in the search, and even if others are later found its primacy may hold weight.

This dissertation takes a fundamentally interdisciplinary approach to cognitive inquiry, building models at the intersection of experimental and longitudinal human subjects studies, computational theory, and linguistic representation. For example, the study of word segmentation builds from the linguistic insight that the syllable is the most prominent metrical unit at the age at which infants begin to segment speech, builds a mechanism to segment syllables using concepts from reinforcement learning, and tests the validity of the mechanism in a performance evaluation and against longitudinal data. The result is a learning model that engages a wide swath of the cognitive science community, from researchers interested primarily in computational learning models to those interested in longitudinal studies of language development.

Another example of the interdisciplinary impact of this thesis is the development in Chapter 5 of a model of analogical change. The core criterion used by the model stems from the serial access mechanism first identified in Chapter 4; a model of how items may be listed in memory is used to develop a criterion for linguistic generalization, and that criterion is then applied to a historical language change. The model identifies both properties of the language undergoing the change and properties of the learner that enabled this change, enabling further study of the relationship between morphophonological structure and language change.

In addition to the specific contributions listed above, a broader interdisciplinary impact of this thesis is the demonstration that extremely simple models can help address long-standing questions of the cognitive structure of language. This work shows that computational modeling need not only be limited to sophisticated statistical learning methods and computationally-intensive approaches. I encourage every cognitive scientist to explicitly model what they study and explore the relationships between their models and human behavior.

References

M. Alegre and P. Gordon. Frequency effects and the representational status of regular inflections. *Journal of Memory and Language*, 40(1):41–61, 1999.

R. H. Baayen, T. Dijkstra, and R. Schreuder. Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language*, 37(1):94–117, 1997.

R. H. Baayen, L. Wurm, and J. Aycock. Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities. *Mental Lexicon*, 2(3):419–463, 2007.

D. Balota, M. Yap, M. Cortese, K. Hutchison, B. Kessler, B. Loftis, J. Neely, D. Nelson, G. Simpson, and R. Treiman. The English Lexicon Project. *Behavior Research Methods*, 39(3):445–459, 2007.

R. Brown. *A First Language: The Early Stages.* Harvard University Press, Cambridge, Massachusetts, USA, 1973.

J. Elphinston. *The Principles of the English Language Digested: or, English Grammar Reduced to Analogy*, volume 1. James Bettenham, 1765.

S. Goldwater, T. L. Griffiths, and M. Johnson. A Bayesian framework for word segmentation: Exploring the effects of context. *Cognition*, 112(1):21–54, 2009.

J. Hay and H. Baayen. Parsing and productivity. In *Yearbook of Morphology 2001*, pages 203–235. Springer, 2002.

M. Johnson and S. Goldwater. Improving nonparameteric Bayesian inference: Experiments on unsupervised word segmentation with adaptor grammars. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 317–325. Association for Computational Linguistics, 2009.

B. MacWhinney. *The CHILDES Project: Tools for analyzing talk.* Lawrence Erlbaum Associates, Mahwah, NJ, US, 2000.

D. Marr. *Vision: A Computational Approach.* Freeman & Co., San Francisco, CA, USA, 1982/2010.

W. Murray and K. Forster. Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review*, 111(3):721–756, 2004.

A. M. Peters. *The Units of Language Acquisition.* Cambridge University Press, 1983.

J. R. Saffran. The use of predictive dependencies in language learning. *Journal of Memory and Language*, 44(4):493–515, 2001.

J. R. Saffran, R. N. Aslin, and E. L. Newport. Statistical learning by 8-month-old infants. *Science*, 274(5294): 1926–1928, 1996.

W. von Humboldt. *On language: The diversity of human language construction and its influence on the mental development of the human species.* Cambridge University Press, 1836/1988. Translated by by P. Heath.

C. Yang. On productivity. *Linguistic Variation Yearbook*, 5:333–370, 2005.

C. D. Yang. Universal grammar, statistics or both? *Trends in Cognitive Sciences*, 8(10):451–456, 2004.