1 At the Intersection of Cognitive Processes and Linguistic Diversity

Andrea D. Sims, Adam Ussishkin, Jeff Parker, and Samantha Wray

Are the dimensions of morphological diversity dependent on the cognitive pathways for processing, storage, and learning of word structure, and if so, how? Conversely, are languages that differ in their morphological structure processed and learned in different ways? This volume examines the relationship between linguistic cognition and the morphological diversity found in the world's languages. As the idea that domain-general cognitive processes and morphological typology are inextricably linked has moved into the mainstream of linguistics, the field has diversified conceptually and methodologically. This introduction to the volume offers an overview of conceptual issues that underpin the volume's papers and some of the methodological trends they reflect. It thus serves as a roadmap for the papers that follow.

1 Introduction

Languages vary substantially with respect to how their morphological systems are structured, including which morphological processes they utilize and how they are used. This volume connects insights from formal theoretical linguistics, psycholinguistics, linguistic typology, and computational linguistics in order to highlight the increasing attention to the ways in which morphology is tied to cognitive processing. As experimental, corpus, and computational methodologies have become more advanced, accessible, and mainstream, recent research has converged on the notion that the cognitive processing of language and typological distributions of morphological structures in the world's languages are intimately linked. In this volume we collect chapters that seek to examine the interaction between linguistic cognition and morphological structures. Our goal is to encourage discussion of questions that cross subdisciplinary boundaries, to highlight the current state of research, and to help shape a research agenda that integrates different methods and approaches.

In addition to this introductory chapter, the volume contains nine chapters that center around three core questions:

1. In what ways is language processing tuned to the morphological structure of a language?

2. What role does cue informativity play in learning and how the lexicon evolves over time?
3. How do system-level principles of morphological organization emerge?

The chapters are grouped into parts defined by these questions. Still, the questions are intertwined and in many cases a chapter whose primary focus relates to one of these questions touches secondarily on one or more of the other questions. Part I contains Chapters 2–4 but see also Chapters 5–7 and Chapter 10. Part II contains Chapters 5–7 but see also Chapter 4. Part III contains Chapters 8–10 but see also Chapters 4–7. The chapters themselves take up different specific questions within these overarching ones, but collectively they offer insight into the relationship between morphological diversity and linguistic cognition. In presenting different approaches side by side, this volume seeks to offer a (partial) view of the landscape of current work.

To this end, this introduction to the volume highlights connections among the volume's chapters. Since readers may not be familiar with all of the methods and models assumed by the individual chapters, we take the opportunity to offer an introduction to broad trends in the field that the chapters reflect. In offering a roadmap for the major issues, we hope to make the volume accessible to readers who might be familiar with the methods and approaches of some chapters but not others, or for whom the questions engaged with here are new.

The remainder of this chapter is structured as follows: Section 2 gives a historical overview of the trajectories of research in psycholinguistics, linguistic theory, and linguistic typology, highlighting the ways in which developments in each field have increased opportunities for productive dialogue and how these developments form the background against which the following chapters are set. One theme that runs through this discussion has to do with the tension between language universals and linguistic diversity. Section 3 summarizes each contribution under the question headings that define the structure of the volume. Section 4 offers some concluding thoughts.

2 A Historical Overview of Research Directions in Psycholinguistics, Linguistic Theory, and Linguistic Typology

In recent years there has been increasing focus in psycholinguistics, linguistic theory, and linguistic typology on the intersection of cognitive processes and linguistic diversity. The fields have followed parallel paths in some respects, as persistent problems in identifying properties shared by all languages have forced all three to reckon with the significance of language-specific differences in grammatical structure. In other respects, however, they have followed different trajectories, arriving at the intersection of cognitive processes and language diversity from opposite directions. While psycholinguistics has always placed the mechanisms involved in language processing and learning at the center of investigation, linguistic theory has long been focused on speakers' structural knowledge of language (their linguistic competence) rather than "performance factors." Linguistic typology has traditionally focused on classifying languages into types and identifying language universals. To the extent that the fields have ended up at the same intersection, it is because the structural traits of languages have become more important to psycholinguistics, on the one hand, and on the other hand, linguistic theory and typology have increasingly turned to language processing and learning to provide explanation for the structural organization of languages and the distribution of linguistic traits in the world's languages. In terms of the present volume, the key observation is that both the parallelisms and the differences in the trajectories of the fields have increased the potential for productive dialogue across them. In the following subsections, we give historical overviews of research directions in psycholinguistics, linguistic theory, and linguistic typology, focusing on the major trends that define the current state of each field and underpin the research presented in subsequent chapters.

2.1 Language Processing, Linguistic Diversity, and Universality

Psycholinguistics takes the modeling of language comprehension, production, and learning as its central goal, but the field in its early days exhibited a kind of language blindness (Bates et al. 2001, Cutler 2009, Norcliffe et al. 2015). A core assumption up to the end of the 1970s was that the language processing system is universal and shared among all humans. Universalist models operated on the premise that processing mechanisms are insensitive to language-specific or item-specific properties. This derived from the observation that children learn whatever languages they are exposed to, suggesting that all humans are born with the same cognitive machinery for language. Inasmuch as the goal was to understand the nature of that machinery, this observation was extended into an assumption, often unstated, that the specific language under investigation was inconsequential to the goal of modeling language processing. (See Norcliffe et al. [2015] for an enlightening discussion and exploration of the history behind this assumption). This resulted in turn in a bias towards studying English and other Western European languages, since these were generally the most accessible to researchers. Only in the 1980s did this core assumption begin to be questioned, with the appearance of experimental results inconsistent with the idea that language processing is universal (e.g., Cutler et al. 1983). Cutler (2009) identifies the rise of questions about how cognitive processing is tuned to the language one speaks (i.e., to the specific properties of particular languages) as the largest paradigm shift in the history of psycholinguistics.

The ways in which the specific grammatical structures of a language may influence how that language is processed and learned have thus become central questions for psycholinguistic investigation. Moreover, when viewed from a crosslinguistic perspective, this shift has profound implications (Frost and Grainger 2000, Marslen-Wilson 2001, Norcliffe et al. 2015). The diversity of grammatical structures found in the world's languages has been exploited to good effect, with studies investigating correlations between the structural properties of a target language and how speakers process or learn that language. Structural differences create naturally occurring experimental conditions that are informative when languages with different structural properties are compared using the same experimental design. This sort of comparison offers insight into the ways in which language processing and learning mechanisms vary depending on the specific properties of a speaker's language, and it becomes possible to disentangle universal patterns in language processing from language-specific ones, revealing how domain-general mechanisms manifest in language-specific ways.

There is now decades' worth of work supporting the conclusion that the processing system is tuned to language-specific structural and distributional properties, although there is not consensus on exactly how. There is not space here to review the rich range of findings (see the overview articles cited above, as well as the chapters in Part I of this volume, inter alia), but morphology is central to this line of work. Lexical access has been shown to be sensitive to the properties of morphological structures on a language-internal basis: prefixedness versus suffixedness (Colé et al. 1989, Hay 2003, Marslen-Wilson et al. 1994), morphological family size (Bertram et al. 2000a, De Jong et al. 2000, Feldman and Pastizzo 2003), degree of semantic transparency (Feldman et al. 2002, Feldman and Pastizzo 2003, Schirmeier et al. 2004, Veríssimo 2018), and productivity (Bertram et al. 1999, 2000c; Lázaro 2012; Wray and Ussishkin this volume), among others. And the crosslinguistic dimension is at play here as well. After all, languages differ substantially in the extent to which they utilize morphological structure. They also differ in how their morphological systems are structured. It thus follows that there might be substantial differences in the cognitive pathways for lexical access across languages.

To take a single illustrative issue, one question that has been the subject of much discussion has to do with whether complex words that are semantically transparent are processed differently from ones that are semantically opaque. As Feldman and Moscoso del Prado Martín (this volume) review, this question has typically been interpreted as a question about morphological decomposition during lexical access: are all words decomposed into morphological constituents at an early stage of processing? Some morphological processing models have posited distinct, cascaded stages of lexical access in which decomposition at the initial stage is entirely form-based. In such a model, early processing is expected to be insensitive to semantic transparency, with both transparent and opaque items subject to decomposition. Data supporting models that are insensitive to semantics in initial stages of processing have sometimes been interpreted through a universalist lens, extending evidence of semantically blind decomposition in one language to a general hypothesis that semantically blind decomposition occurs in all languages. However, crosslinguistic data have cast doubt on this claim, raising questions about how to model differences in experimental results from one language to another, as well as questions about the implications for cognition.

Such questions are commonly approached with behavioral experiments using the lexical decision paradigm. Lexical decision is a word recognition task. A stimulus consists either of a string of letters presented visually on a computer screen or a sequence of speech sounds presented auditorily over headphones. The research participant is asked to respond by quickly judging whether the stimulus is a real word. This deceptively simple methodology has been demonstrated to be quite robust in revealing correlations between behavior (such as participant response time and/or accuracy) and lexical properties of the words themselves, including the token frequencies of the word and its morphological constituents, the productivity of its constituents, and so forth.

To investigate relationships between words or wordforms, the target stimulus may also be preceded by a non-target word (a prime), leveraging the fact that the properties of the prime may affect behavior during exposure to the target. The prime is frequently masked, meaning that it is presented for a sufficiently short duration and with some visual or auditory obfuscation such that the participant is not consciously aware of having seen or heard it. In a commonly used version of the task, the prime shares some morphological structure with the target (e.g., a root). When such a prime facilitates access to the target (i.e., shortens response times), this facilitation is generally interpreted as happening because recognition of both the prime and the target involves accessing a representation for the shared root. The logic is that accessing that representation when processing the prime raises the activation of the root's representation, making it faster to access it again when the target is subsequently observed. Facilitation is thus conventionally interpreted as evidence that morphological decomposition has occurred.

Crosslinguistically, masked priming lexical decision experiments comparing semantically transparent and semantically opaque derived target words have produced a range of results; taken collectively, these have been interpreted as evidence of language-specific tuning. Most notably, in Hebrew, prime–target pairs sharing the same three-consonant root typically facilitate lexical retrieval of the target, even in the absence of semantic relatedness (e.g., *klita* 'absorption' – *taklit* 'a record,' both of which are derived from the root *klt* yet share little if any semantic relationship) in both visual (Frost et al. 1997,

2000) and auditory (Geary and Ussishkin 2019) word recognition. A similar pattern of results has been found in Arabic (Boudelaa and Marslen-Wilson 2000), but these results for Semitic languages contrast with results for English, in which early studies did not find facilitation for semantically opaque prime–target pairs (e.g., *successful–successor*; Marslen-Wilson et al. 1994), and more recent studies have found facilitation but with less robustness (Rastle et al. 2004). In Maltese, a Semitic language whose lexicon has been heavily influenced by Romance languages due to prolonged contact, Semitic pairs that are semantically related and share a consonantal root show priming effects (Ussishkin et al. 2015), whereas pairs analogous to the Hebrew words above (i.e., they share a consonantal root but not semantics) do not show facilitation (Ussishkin 2017). Morphological priming results thus seem to suggest a more central role for morphological structure in lexical access in Hebrew and Arabic, compared to English or Maltese, producing different levels of sensitivity to semantic transparency crosslinguistically; see Feldman and Moscoso del Prado Martín (this volume) for further in-depth discussion of this issue.

Some early models interpreted results of this sort in parameterized terms (terminology from Norcliffe et al. 2015). Parameterized models assume that processing pathways vary from one language to another depending on language type. For example, Marslen-Wilson (2001) posits that this cross-linguistic difference in sensitivity to semantic transparency stems from the non-concatenative nature of Hebrew and Arabic morphology versus the primarily concatenative nature of English morphology. Hebrew words are highly structured according to a system of morphological templates, whereas such a system exists very little for English and only for the Semitic half of the Maltese lexicon. This interpretation is parameterized inasmuch as it assigns causation not directly to speakers' experience with the structural and distributional properties of individual words, but to classifying traits that can be taken as defining the language's 'type' as a whole (concatenative vs. non-concatenative).

While parameterized models are able to handle at least some crosslinguistic variation, they also face a number of conceptual problems. Most notably, evidence of tuning of the processing system at the level of individual affixes or other morphological structures is difficult to accommodate under a strong version of the parameterized hypothesis.[1] Parameterized models likewise appear to be too coarse-grained to account for differences at the level of individual speakers (see Rácz et al. 2015, inter alia). These kinds of problems have

[1] To this we can add that being concatenative or non-concatenative is a property of morphological structures and is not straightforwardly generalizable to languages as a whole. Moreover, in the absence of a specific hypothesis about mechanisms linking specific morphological constructions to language-level effects on language processing, the idea does not move beyond the status of a hypothesis.

motivated a shift towards a more fine-grained approach that focuses on the properties of individual morphological constructions.

Models of morphological acquisition have moved in the same direction. To take a well-known and early example, Pinker (1982) proposed a universal order of acquisition of word order and case morphology: children will produce the dominant word order and use it as a cue in comprehending sentences before they master their language's morphology. However, documentation of actual child language acquisition has shown greater variation in order of acquisition than Pinker's universalist hypothesis predicts. There is evidence that where morphosyntactic case is transparently marked, children attune to case before word order (Slobin 1985, Slobin and Bever 1982). This kind of evidence suggested early on the need for a model of acquisition that is statistically driven by speakers' experience with language.[2] Such a model is further supported by experiments with adults; in a miniature artificial language learning setting, adults are reported to use case marking more often if learning an artificial language with flexible constituent order compared to one with fixed constituent order (Fedzechkina et al. 2017). Ultimately, data suggesting language-specific – even item-by-item – tuning in both processing and learning have fueled an alternative approach which Norcliffe et al. (2015) refer to as experience-based models.[3]

Experience-based models posit that processing and learning are statistically driven and determined by the pathways via which processing has been successful in the past. For example, Plaut and Gonnerman (2000) developed a computational connectionist model of lexical access that produces results that parallel those of masked priming lexical decision studies in Hebrew and English. They show that in their model, in a morphologically rich artificial language (analogous to Hebrew), morphologically related but semantically opaque derived words are primed by their bases, whereas in a morphologically poor artificial language (analogous to English), no priming is found in this condition, although priming is found in semantically more transparent items. In connectionist models, pathways for processing that prove successful are strengthened, making them more likely to be used in the future for processing similar stimuli. Plaut and Gonnerman conclude that the morphologically rich language exhibits priming in the absence of semantic similarity because robust morphological structure (regardless of whether it is dominantly concatenative or non-concatenative) leads to a strengthening of pathways for lexical

[2] For a review of language-specific effects in language acquisition and in language disorders, see Bates et al. (2001).
[3] Norcliffe et al. (2015) categorize syntactic processing models, but their basic division into universalist, parameterized, and experience-based models is also applicable to morphological processing models.

access that take advantage of the informativity of morphological constituents as cues to meaning. In morphologically poor languages this does not happen to the same degree, resulting in a greater reliance on word-level associations between form and meaning.[4] This strengthening of some pathways at the expense of others is experience-based inasmuch as it is a function of the frequency with which speakers have encountered a given (morphological) form and the validity of the form as a cue to meaning among words that the speaker has observed in the past. Plaut and Gonnerman's paper is thus an illustration of how specific language processing mechanisms, with the same initialization conditions, can result in diverse processing pathways when exposed to different input structures. In addition, their chapter illustrates diversification in methodology, which is especially important when considering how best to theoretically model crosslinguistic differences. In this case, experimental methods and computational modeling go hand in hand, with experimental methods providing empirical data and hypotheses about lexical access mechanisms, and computational modeling providing a way to test hypotheses under controlled conditions.

Work of this sort has offered insights into how seemingly conflicting results of word recognition studies in different languages can be reconciled. Feldman and Moscoso del Prado Martín (this volume) use a statistical meta-analysis of English, German, Hebrew, and Maltese data to argue that an experience-based processing model – specifically, one based on discriminative learning (discussed below) – is promising as a way to unify the disparate results of word recognition studies investigating semantic transparency effects. Ultimately, they suggest that uncovering universal aspects of language processing requires more careful attention to the diversity of ways in which words are related to each other. (Since the discriminative framework is non-decompositional, they also call into question the interpretive assumptions made in decompositional models that underpin many universalist and parameterized approaches.)

Within experience-based models, two subtypes can be identified: we term these 'representation-based' models and 'expectation-based' models. In representation-based models, the strength of the memory representation of a target structure determines how a target is processed. Non-interactive parallel dual-route models, also known as "race" models (Baayen et al. 1997,

[4] Here we gloss over the fact that connectionist models tend to be subsymbolic, including Plaut and Gonnerman's. The concepts of 'word' and 'affix' thus have no status as theoretical primitives. However, affixes amount to emergent structures in these models to the extent that they represent reliable form–meaning correspondences that the model can capitalize on for language processing. The question of language-specific tuning can thus be framed as a question about how the processing system identifies the structural units that result in the most reliable form–meaning correspondences for any particular input data.

Baayen and Schreuder 1999, Bertram et al. 2000c), fall here: in this framework, competition between decompositional and whole-word processing routes is mediated by the relative activations of word, stem, and affix lexical entries in a non-interactive way. In turn, the resting activation level of a representation is a function of how successfully it competed in the past. For example, Wray and Ussishkin (this volume) show that in Maltese, the likelihood of a word being parsed during lexical access depends on the productivity of its word class.

Expectation-based models, in contrast, posit that the pathways for language processing are tuned to language structure as a function not just of the representational strength of a given form–meaning pairing, but also as a result of paradigmatic competition among forms as cues to meaning.[5] They thus view representations as elements in a paradigmatically structured network, and questions have to do with the specific mechanisms involved in this paradigmatic competition. Expectation-based models quantify this competition in different ways; one current approach is based on discriminative learning (Baayen et al. 2011, 2019; Baayen and Ramscar 2015; Ramscar et al. 2013). Discriminative learning centers on the idea that learning is sensitive to the informativity of cues. In discriminative learning, form–meaning associations have weights that are modified during the course of learning according to their informativity. While there is not space here to examine its inner workings, discriminative learning offers a specific linking hypothesis about how forms with overlapping distributions compete as cues to the overlapping meaning. The association weight of a form depends on whether it correctly predicts the associated meaning (if not, the association is weakened) and if yes, whether it is the most informative (in which case the association is strengthened). A speaker's prior experience relating cues to meanings thus generates expectations (predictions) about form–meaning relationships, and the cognitive pathways for processing and learning are tuned accordingly. This is the sense in which the model is expectation-based. Discriminative learning has been used to predict reading times (Baayen et al. 2011), visual lexical decision latencies (Feldman and Moscoso del Prado Martín this volume, Filipović Đurđević and Milin 2019, Milin et al. 2017), and child language learning (Ramscar et al. 2013).

Some of the chapters in this volume probe the inner workings of discriminative learning models and its implications for modeling language processing and learning. In addition to Feldman and Moscoso del Prado Martín's chapter, mentioned above, Caballero and Kapatsinski (this volume) examine limitations of discriminative learning and related algorithms. They

[5] The Competition Model (Bates and MacWhinney 1987, MacWhinney 1987) was an early example of this kind of model, focused on language acquisition.

use discriminative learning to identify morphosyntactic and morphosemantic cues in Choguita Rarámuri (an Uto-Aztecan language of Mexico). Unlike other languages to which discriminative learning has been applied, Choguita Rarámuri has a relatively agglutinative structure. Counterintuitive results when applied to this language reveal the "problem of accidentally exceptionless generalizations" (Albright and Hayes 2006; a situation in which a form is unexpectedly learned as a cue to a meaning as a byproduct of the form's rarity) and the "strict teacher problem" (Kruschke 1992; a counterintuitive adjustment of cue weights that results in a form that never occurs in the context of a particular meaning nonetheless being learned as a cue to that meaning). The chapter thus highlights the need to test processing and learning models on a typologically diverse sample of languages.

In a different vein but connecting back to the question of the role of semantic transparency in morphological processing, Needle et al. (this volume) investigate what attributes of English pseudowords (i.e., nonexistent words that are phonotactically possible) affect speakers' acceptability judgments. Among other results, they find that pseudowords that contain at least one existing morpheme, and thus could be decomposed by speakers despite not having clear semantics, are more likely to be accepted as possible words, with some differences in ratings depending on the type of pseudoword (e.g., real root vs. real affix). They explore how the specific mechanisms of two models, a recent multiple-route model (Grainger and Beyersmann 2017) and a discriminative learning model (Milin et al. 2017), might produce this 'shallow parsing' effect. These chapters thus reflect research developments in the direction of comparing specific predictions of experience-based models using wide-ranging methods and data.

In summary, the trend in psycholinguistics has been away from assumptions that all languages are processed in the same way, and towards questions about how the human cognitive system for language processing and learning is shaped by the input that a speaker receives. This, in turn, has naturally fueled investigation of language-specific differences in grammatical structure and thus brought the field to the intersection of cognitive processes and language diversity. As evidence has accumulated for differences in processing and learning from one language to another, but also for differences within a given language according to the structural and distributional properties of a target item, the field has increasingly moved towards experience-based models that are more fine-grained in their predictions about how specific inputs interact with the processing system. Broadly speaking, experience-based models have the potential to accommodate language-specific, item-specific, and speaker-specific differences in the pathways of language processing. Current research (including in this volume) highlights the ways in which the representation of a target structure affects how it is processed, including the ways in which it