HHS Public Access

Author manuscript

Lang Learn. Author manuscript; available in PMC 2017 December 01.

Published in final edited form as:

Lang Learn. 2016 December ; 66(4): 900–944. doi:10.1111/lang.12168.

Learning Additional Languages as Hierarchical Probabilistic Inference: Insights From First Language Processing

Bozena Pajak[a], **Alex B. Fine**[b], **Dave F. Kleinschmidt**[c], and **T. Florian Jaeger**[c]

[a]Duolingo, Inc.

[b]Hebrew University of Jerusalem

[c]University of Rochester

Abstract

We present a framework of second and additional language (L2/L*n*) acquisition motivated by recent work on socio-indexical knowledge in first language (L1) processing. The distribution of linguistic categories covaries with socio-indexical variables (e.g., talker identity, gender, dialects). We summarize evidence that implicit probabilistic knowledge of this covariance is critical to L1 processing, and propose that L2/L*n* learning uses the same type of socio-indexical information to probabilistically infer latent hierarchical structure over previously learned and new languages. This structure guides the acquisition of new languages based on their inferred place within that hierarchy, and is itself continuously revised based on new input from any language. This proposal unifies L1 processing and L2/L*n* acquisition as probabilistic inference under uncertainty over socio-indexical structure. It also offers a new perspective on crosslinguistic influences during L2/L*n* learning, accommodating gradient and continued transfer (both negative and positive) from previously learned to novel languages, and vice versa.

Keywords

second language acquisition; hierarchical probabilistic inference; statistical learning; speech adaptation

Introduction

Infants are born with the ability to learn any of the world's languages. Additional languages can be acquired throughout the life span, but the ability to achieve nativelike proficiency declines with age of first exposure (Hakuta, Bialystok, & Wiley, 2003; Stevens, 1999). What then are the constraints on second and third (or additional) language (L2/L*n*) acquisition in adulthood? One known constraint is that learning new languages as an adult is plagued by negative transfer from the native language (L1), which occurs when the L1 and the target language differ with respect to specific linguistic properties, and the learner incorrectly applies the L1 norm to the L2/L*n*. However, prior native language knowledge has also been

Correspondence concerning this article should be addressed to Bozena Pajak, Duolingo, Inc., 5533 Walnut Street, 3rd floor, Pittsburgh, PA 15232, USA. bozena@duolingo.com.

found to facilitate learning: At least for some grammatical features, learners have an easier time acquiring L2/L*n* properties that already are present in their L1. Standard approaches, from both the emergentist and the nativist traditions, generally agree that L1 knowledge plays an important role in learning subsequent languages (see O'Grady, 2008, Odlin, 2013, and White, 2012, for overviews). Therefore, understanding precisely how and when prior language knowledge leads to interference or facilitation is a pressing question in research on L2/L*n* acquisition.

In this article, we outline a unified framework of both L1 adaptation and L2/L*n* learning as continuous probabilistic inference in response to language input. This framework, we argue, helps reconceptualize the nature of transfer (or crosslinguistic influences) from prior language knowledge. On the one hand, L2/L*n* learning is known to be extremely difficult: Learners struggle with pervasive interference from previously learned languages, and rarely approach native-speaker levels of proficiency. On the other hand, there is a growing literature, as we describe below, demonstrating the astonishing flexibility of adults to learn the statistical properties of languages that they are exposed to in the lab. The theoretical framework we propose brings a new perspective to bear on these seemingly contradictory findings.

At the heart of the proposed framework lie the hypotheses that (a) adult language learners perform continuous probabilistic/statistical inference on their language input, and that (b) this inference process is sensitive to the underlying socio-indexical structure of their linguistic environment, by which we mean talker identity and linguistic generalizations across talkers (e.g., by gender, age, dialect, foreign accent, etc.). The first hypothesis is shared with many previous proposals (discussed below), though, as we argue, some of its consequences are still underappreciated. The second insight—that probabilistic inference and learning should take into account learners' probabilistic, hierarchically structured implicit beliefs about the socio-indexical structure of their linguistic environment—is underexplored in research on L2/L*n* acquisition.[1]

We distinguish variability due to socio-indexical structure from variability due to linguistic context, such as surrounding sound segments or syllable position. Such linguistic context has received comparatively more attention in L1 and L2/L*n* processing and learning (e.g., McMurray & Jongman, 2011; Nearey, 1990, 1997; Nearey & Assmann, 1986; Nearey & Hogan, 1986; Smits, 2001a, b). Here, we are interested in dependencies beyond the linguistic context defined in this sense. Specifically, talkers differ in their realization of phonetic contrasts (e.g., Peterson & Barney, 1952), as they do in their lexical, syntactic, and other preferences (e.g., Weiner & Labov, 1983). Crucially though, talkers tend to not vary randomly. Instead, there is structure in the variability across talkers: Some of the variability across talkers is predicted by talkers' physiological properties (which in turn are correlated with age, gender, etc.) or by their language background (e.g., Great Lakes vs. Texan

[1]Throughout this article, we often use the Bayesian term "belief." For most purposes, belief can be substituted by "knowledge." We use the term belief as it intuitively highlights the uncertainty learners are expected to maintain about their representations of linguistic and socio-indexical structures. Rather than to either know or not know something, learners are taken to hold hypotheses about the structure of language(s) with different degrees of certainty.

American English). This structured variability is what we refer to as *hierarchical indexical structure* (following Kleinschmidt & Jaeger, 2015).[2]

As we describe below, L1 processing requires listeners to overcome—and, in fact draw on—variability between talkers and groups of talkers in order to achieve robust language understanding. We propose that L2/L*n* learning can be seen as an extreme case of the same inference problem. In this view, learning to understand a L2/L*n* constitutes the same fundamental computational problem as adapting to a new L1 dialect or accent. Differences between L1 adaptation and L2/L*n* learning, as well as differences between L2/L*n* learning of different languages, are then primarily attributed to two factors: (a) differences in the strength of the learner's prior beliefs about the L*n* based on previous exposure to other languages (L1 to L*n*–*1*), and (b) the similarity between these prior beliefs and those required to robustly process the L*n*. Two critical contributions of our framework are therefore that (a) it provides a unified view of both L1 processing and L2/L*n* learning as involving the same types of probabilistic inferences and that (b) it helps reconceptualize the nature of transfer in L2/L*n* acquisition by viewing it as learners' inferences about the target language based on their current total language knowledge. This includes rich knowledge about talker- and group-specific distribution of linguistic categories (i.e., knowledge about how linguistic structure is conditioned on socio-indexical structure).

Before launching into the stepwise development of our arguments, we outline our proposal and the structure of the article. The development of our argument falls into three parts. In the first part, we discuss why implicit distributional knowledge of the covariance between linguistic and socio-indexical structure is critical for robust L1 understanding. We then summarize some of the key pieces of evidence that L1 processing, indeed, critically relies on socio-indexical knowledge. With this background established, the second part of our argument turns to L2/L*n* acquisition and to the exposition of the framework we propose. We argue that L2/L*n* learners engage in probabilistic inference over the environment-specific "mini-grammars" they induced for L1 (and other languages previously exposed to), which in turn guides their learning of the target language. Learning a new language thus involves inferring its relationship with previously established patterns. In the final part of our argument, we describe how this reconceptualization of L2/L*n* acquisition naturally captures aspects of L2/L*n* learning that currently lack a unifying explanation. In particular, the proposed framework accounts for the following five well-documented properties of L2/L*n* acquisition: (a) L2/L*n* development is gradual, rather than being limited to an initial transfer from previously acquired languages, and highly variable, as it involves simultaneous maintenance of multiple options for some linguistic properties; (b) transfer can apply from any previously learned language, not only L1; (c) transfer is affected by (actual and perceived) structural similarities between the source language and the target language; (d) transfer is multidirectional in that it can affect previously acquired language knowledge,

[2]It is possible that the brain treats socio-indexical and linguistic context in similar or even identical ways. However, the two types of variability also differ somewhat in the computational challenge they pose for speech perception (see Kleinschmidt & Jaeger, 2015). Depending on the answer to this question, models that were originally intended to capture variability due to linguistic context (e.g., Nearey, 1990; Smits, 2001a, b) might well be extended to capture variability due to socio-indexical structure; indeed, this link was recognized early (Liberman et al., 1967; see Weatherholtz & Jaeger, submitted). Below, we use the term "local environment" to refer to the socio-indexical context, thereby highlighting the potentially qualitative difference between linguistic and socio-indexical context.

including the learner's L1; and (e) transfer involves drawing not only on the specific categories that exist in the source language, but also on the statistical distributions over those categories.

All throughout the article, we illustrate the proposed framework within a normative probabilistic approach that can be naturally interpreted in terms of Bayesian inference. The central ideas behind our proposal are, however, compatible with a few other distributional frameworks, such as, for example, associative learning (e.g., Bates & MacWhinney, 1987; Ellis 2006a, b; MacWhinney, 1983), episodic (Goldinger, 1998) and exemplar-based approaches (Johnson, 1997; Pierrehumbert, 2003; van den Bosch & Daelemans, 2013). We discuss links to and differences from these accounts where appropriate. In developing our proposal, our primary goal is to help readers unfamiliar with this type of framework to develop intuitions about it. We therefore avoid mathematical notation. There are, however, computational implementations of the proposed framework for L1 speech perception (Kleinschmidt & Jaeger, 2015; Nielsen & Wilson, 2008) and L1 sentence processing (Fine, Qian, Jaeger, & Jacobs, 2010; Myslín & Levy, 2016). Detailed development of the formal inference framework applied to L2/L*n* processing can be found in Pajak (2012).

L1 Processing as Hierarchical Probabilistic Inference Under Uncertainty

We begin by introducing two fundamental computational challenges to language understanding: (a) the speech signal is perturbed by noise, causing the mapping between signal and linguistic categories to be nondeterministic, and (b) this nondeterministic mapping varies between talkers. We then review what properties a speech perception system must have in order to achieve robust language understanding despite these two challenges, and what this can tell us about the structure of the implicit linguistic knowledge underlying L1 processing.

Recognition As Inference Under Uncertainty

There is now broad agreement that language comprehension is sensitive to the statistics of the input (for recent reviews, see Kuperberg & Jaeger, 2015; MacDonald, 2013). This sensitivity to linguistic distributions is evident at all levels of linguistic organization. Even the earliest moments of speech processing exhibit sensitivity to implicit knowledge about the distributions of linguistic categories (Feldman, Griffiths, & Morgan, 2009). The recognition of phonological categories and words is similarly sensitive to distributional knowledge (e.g., Bejjanki, Clayards, Knill, & Aslin, 2011; Dahan, Magnuson, & Tanenhaus, 2001; Luce & Pisoni, 1998; McClelland & Elman, 1986; Norris & McQueen, 2008). Beyond word recognition, the incremental integration of information during sentence processing relies heavily on implicit beliefs about lexical and syntactic distributions (e.g., Arai & Keller, 2013; MacDonald, Pearlmutter, & Seidenberg, 1994; McDonald & Shillcock, 2003; Dikker & Pylkkänen, 2013; Tabor, Juliano, & Tanenhaus, 1997; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Trueswell, Tanenhaus, & Kello, 1993).

Drawing on the statistics of the input has, in fact, been shown to be a rational solution to the problem of inferring linguistic categories from the speech signal (e.g., Bejjanki et al., 2011; Feldman et al., 2009; Norris & McQueen, 2008).[3] Even in a cognitively bounded system

that makes rational use of its finite resources (e.g., including time, Griffiths, Vul, & Sanborn, 2012; Lewis, Howes, & Singh, 2014), prediction based on the statistics of the input is a crucial component of language understanding (for discussion, see Kuperberg & Jaeger, 2015). The speech signal is perturbed by noise from multiple sources, including errors during speech planning, muscle noise during production, ambient noise from the environment, and noisy neuronal responses in the perceptual system. Although these types of noise differ in many important aspects, they have a common consequence: Noise makes the mapping between linguistic categories and the acoustic signal nondeterministic and, thus, the inverse mapping from the signal to the categories is also nondeterministic. This makes the recognition of linguistic categories—and language understanding more generally—a problem of *inference under uncertainty*.

Specifically, each linguistic category can be thought of as a probability distribution, a function specifying how likely each possible cue value is, given a particular category. The rational solution to the problem of recognizing phonological categories—as examples of linguistic categories—relies on knowledge of these distributions. Bayes' rule describes the exact relationship between the cue distributions and the categorization function of a rational listener. Figure 1 depicts this for the relation between voice onset time (VOT)—one of the primary cues to voicing in English—and the phonological categories /b/ and /p/. The classification function predicted by Bayes' rule, as shown in Figure 1, provides a good qualitative and quantitative fit against human behavior in phonetic categorization tasks (e.g., Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Kleinschmidt & Jaeger, 2015).

The problem of inference under uncertainty is not limited to the recognition of phonological categories, but extends across all levels of linguistic organization. Although many important questions remain about the mechanisms that underlie such inferences, rational models have been found to provide good qualitative and quantitative fits against human language processing at these higher levels of linguistic organization as well (e.g., Boston, Hale, Kliegl, Patil, & Vasishth, 2008; Demberg & Keller, 2008; Norris & McQueen, 2008; Smith & Levy, 2013; for further references, see Kuperberg & Jaeger, 2015). Beyond robustly inferring the intended message from noisy input, implicit probabilistic knowledge can also increase processing speed, for instance, through efficient allocation of attentional resources (Smith & Levy 2013).

In summary, there is converging evidence that (a) the computational systems underlying language comprehension involve implicit probabilistic knowledge about the statistical distributions of linguistic categories, and that (b) this knowledge plays a crucial role in language understanding. However, as we discuss next, reliance on implicit probabilistic knowledge is only beneficial to the extent that this knowledge reflects the actual statistics of linguistics distributions. This turns out to be critical, as the probabilistic mapping between the signal and linguistic categories is variable, changing depending on the local environment.

[3]Rational here is to be understood in the sense of Anderson (1990). A rational solution is one that makes optimal use of available information.

Variability in Mapping Between Signal and Linguistic Categories

Linguistic distributions change depending on the talker, genre, and other socio-indexical variables. This makes linguistic distributions nonstationary, at least from the perspective of language users. In research on speech perception, this problem is known as lack of invariance although this term was originally used to refer to variability in linguistic distributions due to linguistic (rather than socio-indexical) context, such as differences in the realization of onset consonants depending on the following vowel (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; see also Nearey, 1990; Smits, 2001a, b). Different talkers produce instances of the same category differently, using different acoustic-phonetic cues or cue values (e.g., Allen, Miller, & DeSteno, 2003; McMurray & Jongman, 2011; Newman, Clouse, & Burnham, 2001). Figure 2 illustrates this for the VOT example from Figure 1 (for further examples and discussion, see Weatherholtz & Jaeger, submitted).

As can be seen in Figure 2, the rational solution discussed in the previous section is only rational as long as the listener makes the correct assumption about the mapping between acoustic-phonetic cues and linguistic categories. If a listener assumes that the probabilistic mapping between signal and linguistic categories is stationary, this will systematically and negatively affect language understanding. Imagine, for example, a listener with the implicit probabilistic beliefs corresponding to the solid blue line in Figure 2. If that listener receives input from a talker, who produces /b/ and /p/ according to the distributions corresponding to the dashed orange line in Figure 2, the listener will frequently hear /p/, when the talker in fact intended to produce a /b/.

Between-talker variability thus has two immediate consequences. First, listeners might need to adapt whatever implicit phonetic beliefs they hold when they encounter a novel talker that deviates from previously encountered talkers. We can think of this as learning a language model, specifying a set of probabilistic mappings between the signal and linguistic categories for the novel talker—essentially, a probabilistic mini-grammar for that particular talker (Kleinschmidt & Jaeger, 2015). And second, even if a particular talker has previously been encountered, listeners are never quite certain which previously learned language model is appropriate in the current circumstances. Put differently, between-talker variability makes language understanding a problem of inference under uncertainty not only about linguistic categories, but also about the appropriate language model for the current local environment. The consequences of between-talker variability are not limited to speech perception (although they are perhaps starkest in this domain). Rather, the logic outlined above for speech perception extends to lexical and syntactic processing: Reliance on implicit knowledge of linguistic distribution only facilitates efficient sentence processing if language users' implicit beliefs sufficiently closely reflect the actual statistics of the current local environment (see Fine, Jaeger, Farmer, & Qian, 2013; Myslín & Levy, 2016; Yildirim, Degen, Tanenhaus, & Jaeger, 2015).

Overcoming Variability: Evidence from L1 Processing

Now that we have established the conceptual framework of inference under uncertainty about both linguistic categories and the appropriate language model for the current local environment, we summarize some of the key findings from research on L1 language

processing that illustrate how listeners overcome the challenge raised by between-talker variability. We split this summary into two sections, corresponding to the two consequences of variability introduced above. This will establish the conceptual framework that we then extend to L2/L*n* learning.

Learning Between-Talker Variability

Imagine a situation in which a listener encounters a novel talker whose acoustic realizations of linguistic categories (e.g., her pronunciations) deviate from previously encountered talkers. In this situation, listeners need to adapt their implicit beliefs about linguistic distributions for the current environment.[4] Indeed, a growing body of work suggests that L1 speech perception in such situations relies on continuous, implicit statistical learning. In situations with which they have little prior experience, listeners appear to rapidly adapt to the statistics of the acoustic cues associated with different phonetic categories. The main source of evidence for this comes from phonetic recalibration (or phonetic perceptual learning) studies, where listeners hear a sound that is acoustically ambiguous between, say, /b/ and /p/. If a listener hears this sound in a context which implies that it was intended to be a /b/ (e.g., a word that can end in /b/ but not /p/, like *stub*), then they will "recalibrate" their /b/ category, classifying more sounds on a [b]-to-[p] continuum as /b/ after exposure (e.g., Bertelson, Vroomen, de Gelder, 2003; Eisner & McQueen, 2006; Kraljic & Samuel, 2005; Norris, McQueen, & Cutler, 2003; for further references, see Kleinschmidt & Jaeger, 2015).

There are two reasons to think that this adaptation is a form of probabilistic inference. First, as listeners in perceptual recalibration experiments are exposed to more and more evidence from a particular talker, their behavior gradually changes in ways predicted both qualitatively and quantitatively by rational inference under uncertainty about the mapping between linguistics cues and categories (Clayards et al., 2008; Kleinschmidt & Jaeger, 2011, 2012, 2015). The type of learning behavior that such a model predicts is illustrated schematically in Figure 3.

Second, listeners seem to adapt not just to differences in the mean cue values for a category, but also the variance of these category-specific cue distributions (e.g., Bejjanki et al., 2011; Clayards et al., 2008; Kleinschmidt & Jaeger, 2012; for further discussion, see Kleinschmidt & Jaeger, 2015). This follows readily under a rational inference account of between-talker variability, in which adaptation results in changes to listeners' probabilistic beliefs about the shape of the relevant distributions, including their variance (see Kleinschmidt & Jaeger, 2015). Although questions remain about the precise mechanisms, it is now clear that adaptation also occurs in more complex pronunciation shifts, for example, when encountering a dialect- or foreign-accented talker (Baese-Berk, Bradlow, & Wright, 2013; Bradlow & Bent, 2008; Weatherholtz, 2015; but see Best et al., 2015, for limitations). Further, there is evidence that adaptation is not just specific to the linguistic input that has

[4]Some between-talker variability might be dealt with by listener's prelinguistic perceptual normalization (for references and discussion, see Weatherholtz & Jaeger, submitted). However, such normalization is insufficient to account for all systematic variability between talkers (Johnson, 2005). Instead, some variability is idiolect-, sociolect-, or dialect-specific and has to be learned on a talker-by-talker basis (e.g., Johnson, 2005, Pierrehumbert, 2003).

been observed from a talker. Rather, adaptation can generalize to other sounds (Kraljic & Samuel, 2006) and words (Maye, Aslin, & Tanenhaus, 2008; McQueen, Cutler, & Norris, 2006; Weatherholtz, 2015) not heard previously from the novel talker.

Similar adaptation to novel talkers has been observed for deviation from previously encountered phonotactics (Kraljic, Brennan, & Samuel, 2008), prosody (Kurumada, Brown, Bibyk, Pontillo, & Tanenhaus, 2014), lexical usage (e.g., Metzing & Brennan, 2003; Creel, Aslin, & Tanenhaus, 2008; Grodner & Sedivy, 2011; Yildirim et al., 2015), and even syntactic distributions (Fine et al., 2013; Farmer, Fine, Yan, Cheimariou, & Jaeger, 2014; Farmer, Monaghan, Misyak, & Christiansen, 2011; Hanulikova, Van Alphen, Van Goch, & Weber, 2012; Kamide, 2012). For example, Fine et al. (2013) demonstrated that listeners can rapidly and implicitly learn the statistics of a novel local environment. Participants read sentences that had either a matrix verb or relative clause structure, as illustrated in the following two examples:

The experienced soldiers warned about the dangers...

a. before the midnight raid. (*warned* as a matrix verb)

b. conducted the midnight raid. (*warned* as a participle in a relative clause)

At *warned about the dangers*, these sentences are temporarily ambiguous: Participants so far do not know whether the sentence they are reading will have the structure in (a) or in (b). This ambiguity is resolved at the underlined material in (a) and (b), allowing participants to discover the structure of the sentence they are reading. Therefore, reading times at the disambiguating region (underlined in the example) provide an index of how unexpected the observed structure was for subjects. Indeed, reading times at disambiguation are higher for subjectively less probable structures (in this case, relative clauses) than for more probable structures (here, matrix verbs; e.g., MacDonald, Just, & Carpenter, 1992).

If listeners are adapting to the distribution of main verbs and relative clauses in the local environment, their implicit beliefs about these probabilities should change. This change should be reflected in changes in the reading times for the disambiguation region. This is indeed what Fine et al. (2013) found. For example, when relative clauses were locally highly probable, subjects became better at reading relative clause sentences, and worse at reading main verb sentences. In fact, fewer than thirty relative clauses were necessary to override the expectation for matrix verbs. Evidence that these changes in reading times indeed reflect changes in probabilistic beliefs about the distribution of syntactic structures comes from anticipatory eye-movements during spoken language understanding (Kamide, 2012) and from event-related potentials (Hanulikova et al., 2012). Related modeling work by Fine and colleagues suggests that syntactic adaptation of this kind can be successfully captured using the same Bayesian approach described above for speech perception (Fine et al., 2010; Kleinschmidt, Fine, & Jaeger, 2012).

In summary, research on L1 processing suggests that listeners can learn the statistics of novel local environments (e.g., a novel talker). The evidence summarized so far leaves open whether listeners have a single language model that they continuously adapt to adequately reflect the statistics of their recent experience, readapting every time these statistics change.

As we discuss next, this does not seem to be the case. Rather, there is evidence that listeners can represent several different language models as part of their implicit L1 knowledge.

Representing Between-Talker Variability

A substantial part of the variability in the linguistic signal is systematic—it is predictable based on socio-indexical variables like talker identity, sociolect, dialect, accent, etc. A comprehension system that merely relies on continuous adaptation would fail to take advantage of this structure. Instead, a rational solution to a world in which listeners encounter the same talker repeatedly is to remember what one has learned about that talker (see Kleinschmidt & Jaeger, 2015). Further, a rational listener should aim to learn generalization over similar previously encountered talkers, allowing the listener to more effectively adapt to novel talkers based on similar previous experiences. In short, a rational listener should represent knowledge about the covariation between linguistic features and socio-indexical features (e.g., talker identity or talker groups), thereby capturing the systematic aspects of between-talker variability. This idea is illustrated in Figure 4, where each node corresponds to a language model (or mini-grammar) for a particular talker (terminal nodes) or group of talkers. It is in this sense that a rational listener is expected to have rich beliefs about the socio-indexical structure underlying the linguistic signal.[5]

Indeed, research on speech perception provides compelling evidence in support of this view. The most basic evidence comes from studies that have found adaptation to a novel talker to persist over time, even after listeners are exposed to other talkers. For example, Eisner and McQueen (2006) had participants adapt to a novel talker and then tested them either immediately after exposure or with a 12-hour delay. Although the latter group of participants left the lab and received input from other talkers, Eisner and McQueen found no difference in the strength of talker-specific adaptation between the two participant groups (see also Goldinger, 1996). Similar evidence is beginning to emerge for sentence processing (Wells, Christiansen, Race, Acheson, & MacDonald, 2009).

There is also evidence that listeners form novel generalizations across talkers, for instance, based on dialect- or foreign-accented speech (Baese-Berk et al., 2013; Bradlow & Bent, 2008; Weatherholtz, 2015). Critically, listeners draw on these generalizations during speech perception (e.g., Johnson, Strand, & D'Imperio, 1999; Niedzielski, 1999; Strand, 1999; Walker & Hay, 2011). For example, listeners' interpretation of the very same acoustic information is affected by top-down information about the group membership of the talker who produced it (e.g., a male or female face, Johnson et al., 1999; Strand, 1999; being informed that a talker is from "Canada" or "Detroit," Niedzielski, 1999). Evidence of similar

[5]There are other models that can account for listeners' sensitivity to some socio-indexical variables. For instance, episodic models—where speech recognition is mediated by detailed acoustic traces of each word token ever heard (e.g., Goldinger, 1998; Johnson 1997; Pierrehumbert, 2003)—can account for learning and sensitivity to socio-indexical variables like talker identity. By storing each word as it is perceived, information about the talker's identity is encoded implicitly in the detailed acoustic features of the word, and any unusual pronunciations are stored directly. However, existing episodic models struggle with generalization to unheard words (Cutler, Eisner, McQueen, & Norris, 2010), or to groups of talkers without additional abstraction. It is possible to extend these models by adding such abstraction, for instance, in the form of storing episodes at sublexical, phonetic-category-sized granularity, or "tagging" exemplars with socio-indexical variables (Johnson, 2013), and this moves them towards implementing the sort of computations we propose, that is, tracking the talker- or group-specific distributions of cues for each phonetic category (see Kleinschmidt & Jaeger, 2015).

generalizations based on socio-indexical structure is beginning to emerge for phonotactic (Staum Casasanto, 2008), lexical (Walker & Hay, 2011), pragmatic (Kurumada, 2013), and syntactic processing (Hanulikova et al., 2012).

While it remains an open question how exactly listeners represent socio-indexical structure, findings like these suggest that even L1 knowledge involves rich implicit beliefs about the socio-indexical structure that underlies between-talker variability. Listeners do not just adapt their language models to novel talkers. They also represent these novel models, form generalization across them, and draw on this knowledge to facilitate language understanding. As a consequence, even a monolingual listener, when first exposed to a novel L2, already has implicit beliefs about the way in which talkers differ from each other. Overall, these implicit beliefs about the structure of the world are advantageous: They allow recognition of previously encountered talkers (rather than learning from scratch) and efficient generalization to similar talkers (rather than treating all novel talkers as the same). In Bayesian terms, strong prior beliefs about what types of talkers there are in the world mean that listeners need less evidence from a novel talker to determine what type of language model will be adequate. This in turn will mean that the language model used by the listener will more quickly reflect the actual statistics of the talker (see Figure 3), reducing the risk of misrecognition (see Figure 2).

However, with strong prior beliefs about the way in which talkers vary, there is also a price to pay: When confronted with a novel talker that does not follow any previously encountered pattern, adaptation becomes harder. This is essentially a consequence of rational inference under uncertainty. In order to deal with the noisy signal, which creates uncertainty, listeners combine the bottom-up input with their prior beliefs; this means that prior beliefs can change what listeners perceive (e.g., Feldman et al., 2009). When prior beliefs are particularly strong, they can therefore be difficult to overcome. As we discuss next, this logic extends to L2/L*n* learning.

L2/L*n* Acquisition as Hierarchical Probabilistic Inference Under Uncertainty

Thus far, we have argued that L1 speaker knowledge is best understood as a set of language models (or mini-grammars) that encode the hierarchical structure of the listener's linguistic environment and that are continuously being adapted to incoming input. In this section, we extend this architecture to L2/L*n* learning. We argue that a multilingual learner's linguistic knowledge can be characterized as a set of grammars that, similarly, capture the hierarchical indexical structure of the linguistic environment and are continuously being adapted in response to input from the additional languages being learned. This proposal views L2/L*n* learning as in some sense an extreme version of the type of adaptation that even L1 users need to master in order to overcome dialect, sociolect, and individual differences in pronunciation, as well as other linguistic variation. Within this framework, then, differences in learners' ability to acquire additional languages and the ability to adapt to new language properties (as well as general limitations in the ability to learn) are at least to some extent a function of the amount of accumulated knowledge that provides learners with strong biases about how to interpret the incoming input. We begin with the critical assumptions that underlie the proposed framework: (a) adults are able to perform implicit probabilistic

analyses on nonnative language input, (b) one of the main sources of limitations on L2/L*n* acquisition is the learner's prior language background, and (c) the bilingual or multilingual environment of a language learner can be characterized as an extension of hierarchically structured variability within L1.

Statistical Learning in L2/L*n* Acquisition

The justification for assuming adult sensitivity to statistical cues comes not only from the work on L1 processing and adaptation we discussed earlier, but also from a growing body of work on adult language learning (see Rebuschat, 2015). Adults have been shown to attend to statistical cues when learning novel phonetic categories (e.g., Lim & Holt, 2011; Pajak & Levy, 2011; Wanrooij, Escudero, & Raijmakers, 2013), word boundaries (Endress & Mehler, 2009; Saffran, Newport, & Aslin, 1996), phonotactics (Onishi, Chambers, & Fisher, 2002), grammatical categories and dependencies (Reeder, Newport, & Aslin, 2013), as well as morphosyntactic and syntactic structure (Fedzechkina, Jaeger, & Newport, 2012; Hudson Kam, 2009; Wonnacott, Newport, & Tanenhaus, 2008). Adult sensitivity to statistical cues has not only been demonstrated in learning a single new language, but also in tracking the statistics of multiple languages within a single laboratory session (Gebhart, Aslin, & Newport, 2009; Weiss, Gerfen, & Mitchel, 2009).

Questions about the role of statistical learning in L2/L*n* acquisition do, however, remain. First, it is still largely an open question whether statistical learning persists long enough to subserve L2/L*n* acquisition. While some recent studies have found effects of distributional training to persist for months even after relatively brief exposure (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Escudero & Williams, 2014), more work is needed to establish what type of short-term statistical learning translates into long-lasting L2/L*n* knowledge. Second, adults are known to have more difficulty than infants in attending to certain statistical properties of a new language. A well-known example is that of L1-Japanese L2-English learners, who have extreme difficulty learning the /r/-/l/ distinction, both in perception and production (e.g., Miyawaki, Strange, Verbrugge, Liberman, Jenkins, & Fujimura, 1975). Similarly, adults appear to fail in some laboratory tasks, for example, when learning some L2 phonetic categories from statistical cues alone (e.g., Goudbeek, Cutler, & Smits, 2008), when learning certain word orders in an artificial language (Culbertson, Smolensky, & Legendre, 2012), or in some cases of segmenting words from a continuous speech stream (Finn & Hudson Kam, 2008; Newport & Aslin, 2004). However, despite the above findings, we argue that learners are on average striving to be rational and that at least some of these apparent failures of adult learners to successfully infer linguistic categories from statistical cues are in fact not convincing counterexamples to this claim. On the contrary, such counterexamples can be explained by the proposed framework, as long as we keep in mind that the probabilistic inferences learners need to conduct are limited by their cognitive resources.

Sources of Limitations in L2/L*n* Acquisition

Achieving nativelike proficiency in a nonnative language is extremely rare, and certain errors tend to persist regardless of the amount of exposure, especially in the domain of phonology (e.g., Han, 2004). Why is this the case, and how is it compatible with the

approach we are advocating? Many researchers attribute the difficulty of L2/L*n* learning relative to L1 acquisition to maturational factors (e.g., Abrahamsson & Hyltenstam, 2008 Johnson & Newport, 1989). However, there is also evidence that neural plasticity for language learning is not completely lost in adulthood, and nativelike attainment in L2/L*n* acquisition might be possible (see Birdsong, 2009; Moyer, 2014). Some have argued that the apparent limitations of L2/L*n* learning might at least in part be due to differences in incentive and the time dedicated to the learning between infants acquiring their native language(s) and the typical adult L2/L*n* learner (e.g., Marinova-Todd, Marshall, & Snow, 2000). Others have argued that foreign accents and other apparent failures to converge against nativelike proficiency in speech production could be at least in part a consequence of encoding one's social identity (Gatbonton, Trofimovich, & Magid, 2005; Moyer, 2007). These arguments do not necessarily call into question that L2/L*n* acquisition is difficult, but they challenge the assumption that all deviations from the target L2/L*n* are due to an inability to fully acquire the new language.

To the extent that the factors such as motivation or social identity do not explain all the challenges and limitations in L2/L*n* learning, we believe that many of the learning difficulties follow naturally from the hierarchical inference framework that we propose here. In this framework, L2/L*n* learners implicitly strive to behave rationally given the total knowledge they currently possess. In particular, learners' previously acquired language knowledge constitutes strong implicit prior beliefs about the new target language. This prior knowledge contains useful information that allows learners to make fairly accurate implicit guesses about many properties of the target language. At the same time, however, this prior knowledge can also hinder learning, or even prevent learners from attaining a native-speaker level of proficiency. This does not mean that learners on average are not behaving rationally; it simply means that they are trying to take advantage of their prior knowledge, which in some cases leads them astray.

How are the limitations on L2/L*n* learning compatible with listeners' often rapid and seemingly effortless adaptation to the properties of L1 speech? In fact, even in adaptation to novel L1 properties (e.g., accented speech), we can sometimes observe the pervasive influence of L1-based prior beliefs. For example, Idemaru and Holt (2011) showed that while listeners adjust their speech categorization after hearing only five instances of an accented word, this kind of statistical learning quickly asymptotes. Even after five consecutive days of exposure to accented speech, listeners' categorization responses did not reflect the underlying sound distribution, but rather remained intermediate between their long-term L1 representations and the target accent. This demonstrates that learners' prior language knowledge strongly guides (but therefore also constrains) adaptation even in L1 use, to the point that prior knowledge can even block full adaptation.

Given results like these, it is only natural to expect that prior language knowledge may be strong enough to interfere with statistical learning of any additional language, by which we mean a biasing role of previously learned language(s) when implicitly inferring the underlying structure of the new language. Such blocking of statistical learning in L2 has in fact been modeled computationally. For example, McClelland, Thomas, McCandliss, and Fiez (1999) showed that the inability of L1-Japanese speakers to perceptually separate the

English /r/ and /l/ categories naturally falls out of assuming the well-established representations of the relevant phonetic category distributions in Japanese, thus demonstrating computational validity of this explanation, which had previously been offered by many others (e.g., Miyawaki et al., 1975; for a related approach and the idea of L1 neural entrenchment, see MacWhinney, 2012). This means that at least some failures to converge against native proficiency may be best understood as the price that language learners pay for an efficient learning system—a system in which the search through a vast hypothesis space (to determine a grammar for a new language) is made more feasible by relying on prior implicit beliefs about how language is structured. Similar points are made by Ellis (2006a, b), who discusses how apparent irrationalities of L2 acquisition follow from principles of associative learning, or Flege (1999), who notes how foreign accents may arise “not because one has lost the ability to learn to pronounce, but because one has learned to pronounce the L1 so well” (p. 125).

In this context, it is noteworthy that the L1 bias can—under some circumstances and at least to some degree—be overcome, thus suggesting that learners’ difficulties are not all due to an intrinsic inability to learn some properties of a new language. The case of /r/-/l/ learning by L1-Japanese speakers is a canonical example of the difficulty of L2 acquisition. Yet improved learning has been shown even in this difficult case, as long as the learners were provided with stronger support for distributional learning: either through adding more variability to signal irrelevant phonetic dimensions (e.g., Lim & Holt, 2011; Kondaurova & Francis, 2010) or by exaggerating the natural distributions until some initial learning has taken place (e.g., Escudero Benders, & Wanrooij 2011; Kondaurova & Francis, 2010). Based on these results, new L2 linguistic structures will only be induced when the observed signal is sufficiently improbable (and thus unexpected) under the old L1 language model. The limitations on L2/L*n* acquisition do not, therefore, argue against learners’ striving to be rational. Some of these limitations are, in fact, the best possible outcomes given the profound influence of prior language knowledge.

Hierarchical Indexical Structure of a Multilingual Linguistic Environment

The linguistic environment of a multilingual learner is well captured with the kind of hierarchical indexical structure that, as we have proposed, characterizes the environment of a monolingual speaker. For a monolingual speaker, the structure includes clusters of talkers, dialects, and so on (cf. Figure 4). For a multilingual speaker, on the other hand, the structure is far more complex. It includes multiple different languages, where each language has its own internal structure, as illustrated in Figure 5.

From a typological perspective, languages naturally cluster in terms of their similarity. For example, in the hypothetical scenario illustrated in Figure 5, the linguistic environment might include two groups of languages, such as Germanic (G_1) and Romance (G_2), where the Romance group splits further into West-Romance and East-Romance. It is in principle possible to find an objective grouping of languages for any multilingual environment. However, this objective grouping might differ from how the learner actually perceives and represents languages, as we discuss in more detail in the next section. Critically, the proposed hierarchical inference framework is based on the idea that learners are able to

represent in some way this socio-indexical structure of their linguistic environment, although the perceived structure will deviate from the actual structure throughout L*n* acquisition.

The Hierarchical Inference Framework in Multilingual Learning

After having discussed the three critical assumptions that underlie the proposed framework, we elaborate on our proposal that L2/L*n* learners engage in hierarchical probabilistic inference. In particular, we discuss two important properties of the framework. First, learning occurs hierarchically: The learner makes simultaneous (largely implicit) inductive inferences not only about the properties of the target language, but also about the higher-level structure of those properties. This includes assessing the overall similarities and differences between languages in order to assign them to appropriate clusters, as well as tracking the properties shared by all languages. These inferences rely on continuous, implicit statistical learning, which allows learners to keep adjusting their implicit beliefs as a function of received language input. Second, learners' inferences are probabilistic, which means that learners maintain implicit beliefs about different possible language models, where each model is associated with a certain degree of uncertainty, as reviewed for L1 earlier.

An example of a hypothetical multilingual listener's structured beliefs is shown in Figure 6, where L_{any} represents "any language" that encompasses all languages in the hierarchy (Pajak, 2012). It is the abstract knowledge that emerges from all previously learned languages, capturing the learner's implicit beliefs as to what a generic language might look like. L_{any} is related to the traditional concept of interlanguage (Selinker, 1972, 1992); the crucial difference is that L_{any} is not a representation of any particular language, but rather the knowledge that emerges from all previously learned languages. The L_{any} proposal parallels what we have proposed for the organization of L1 knowledge, where higher-level nodes are distributions over the properties of individual speakers, groups of speakers, dialects, and so on (see Figure 4). When considering the case of learning multiple languages, we build additional structure on top of the structured representations of an individual's L1.[6]

The inferred clusters in the hierarchy reflect the perceived structural similarities between the languages. The closer two languages are in the inferred structure, the stronger the learner's implicit beliefs that they share many properties. For an ideal learner, the inferred structure would correspond to the objective typological similarities between languages. For actual learners, however, the perceived similarities between languages will be distorted. In particular, learners may view languages as more similar due to learning them under similar circumstances (e.g., classroom instruction), or due to top-down beliefs about language relatedness. Furthermore, these inferences are also modulated by the degree of uncertainty about previously learned languages, which is in turn determined by language proficiency, recency and regularity of use, and so on (see also Rothman, 2015, for a discussion of the

[6]For a monolingual speaker, L_{any} representations would be predominantly influenced by L1, but would not be equal to L1 representations. L_{any} captures learners' guesses about a generic language, and these guesses will necessarily include some properties distinct from L1, such as an expectation that languages differ in their lexicons, sound inventories, and so on, which are possibly influenced by top-down knowledge about the possible and likely shapes of grammars. These representations may arise from the simple realization that there exist languages other than the learner's L1, or from contact with nonnative speakers, among other factors. What exactly such L_{any} representations for a monolingual speaker look like is an empirical question that we leave for future work.

factors that might be involved in how L3/L*n* learners implicitly assess between-language similarity). The role of these additional factors is expected to be particularly prominent in the initial stages of acquisition, when the evidence from the target language input is limited. Later on we discuss how these aspects of the framework relate to empirical findings in L2/L*n* acquisition.

Most critically, the hierarchical inference framework redefines the concept of language transfer. Instead of viewing it as a direct transfer of properties from a known language to the target language at the outset of acquisition, crosslinguistic influences occur in this framework indirectly via L_{any}, as well as any other intermediate clusters of languages. In many other models, learners are assumed to begin the acquisition of a language by copying all the properties of another known language (see White, 2015, for an account from the Universal Grammar perspective and MacWhinney, 2012, from an emergentist perspective). In our framework, the initial state of any L*n* is viewed not as the properties directly transferred from previously known languages, but rather as sets of hypotheses about the L*n* grammar. These hypotheses, which are the hierarchically structured, implicit probabilistic beliefs arising from experience with previously learned languages, guide learners' best guesses about what the new language's underlying grammar might look like.[7] In other words, these hypotheses are the possible language models that the learner entertains at the outset of acquisition, and they include the learner's guesses about new language's place in the inferred hierarchy. The hypotheses might be based on (a) the learner's implicit prior beliefs about the specific properties of any previously learned language, (b) the learner's inferences about L_{any}, (c) the learner's top-down beliefs, if any, about the relationship of the target language to the known languages, and (d) any learning biases. According to this framework, then, so-called transfer from previously learned languages is observed because, when learners posit that the L*n* is part of a given language cluster, they assume that it shares some properties with other languages in that cluster.

Hierarchical Probabilistic Inference and L2/L*n* Learning Data

In this section, we articulate specific predictions that follow from the hierarchical inference framework and discuss them in light of empirical findings in different areas and aspects of L2/L*n* acquisition. We structure our discussion around five well-known properties of L2/L*n* acquisition and crosslinguistic influences.

L2/L*n* Development is Gradual and Variable

In the hierarchical inference framework, L2/L*n* development is characterized by slow changes to the learner's implicit beliefs about the target language. Learners begin with a set of hypotheses about the target language that are largely based on their prior beliefs about previously learned languages, and then gradually adjust those hypotheses as they obtain more input from the target language. Given that learners continuously entertain multiple possibilities for the underlying language model, each with a different amount of uncertainty,

[7]Note that this way of looking at between-language transfer is very similar to how transfer of knowledge is understood in hierarchical Bayesian inference (see Qian, Jaeger, & Aslin, 2012). Learners are assumed to form hierarchically structured representations, which then facilitate both the formation of abstract rules and principles, and their transfer to novel problems and environments.

we expect to observe large variability in a beginning learner's production and comprehension of the target language. For example, learners might accept two possible word orders for a given structure: one that is consistent with the L*n* input they received and another that is consistent with the equivalent word order in their L1. As learners receive more input from the target language, and thus accumulate more evidence for the targetlike properties, they are expected to gradually transition to relying more on their observations in the target language relative to their prior knowledge. This means that we expect gradual changes in learners' beliefs about the L*n* grammar, as reflected in their language production and comprehension, slowly reducing the influence of other known languages.

In standard linguistic formalist approaches, transfer from L1 is assumed to occur only at the onset of L2 acquisition, and subsequent learning consists of stages during which the initial grammar is molded into a shape approaching the target grammar (for overviews, see White, 2009, 2015). Within these approaches, the influence of prior language knowledge is thus a part of L*n* acquisition only to the extent that learners make use of the properties transferred at the beginning of learning. Furthermore, there is no expectation of gradual changes in the influence of previous language knowledge, as L*n* acquisition is assumed to proceed in stages. Recently, several researchers have criticized these approaches for ignoring the gradience and variability in L2 development, offering new proposals that allowed for "optionality" in the grammars of learners throughout L2 acquisition (e.g., Multiple Grammars Theory, Amaral & Roeper, 2014; or Modular On-line Growth and Use of Language, Sharwood Smith & Truscott, 2014).

We believe that the hierarchical inference framework is a better response to the empirical reality of gradual development than optionality. Indeed, evidence increasingly points to a continuous development in L2/L*n* acquisition that is characterized not only by gradual changes, but also by large variability in using targetlike and other-known-language-like elements (e.g., Amaral & Roeper, 2014; Wunder, 2011). This variability persists across acquisition: from beginning learners (e.g., Rothman & Cabrelli Amaro, 2010) to advanced L2/L*n* users (e.g., Papp, 2000), and what changes across proficiency levels is the frequency with which different options are produced. This is exactly what falls out of the postulates of the hierarchical inference framework.

Relatedly, it has been found that the relative frequency of producing alternative structures in a new language (e.g., expressing vs. dropping a subject pronoun) is affected by the number of previously learned languages that use those structures (De Angelis, 2005). For example, L1-Spanish intermediate learners of Italian—where, as in Spanish, subject pronouns are optional—produce a higher rate of subject pronouns in Italian if they had previously learned two obligatory-subject languages (L2-English, L3-French) relative to the case of having learned only one such language (L2-English). Intuitively, this seems to suggest that learners take individual languages as evidence, based on which they draw inferences about new languages—an idea that is inherent to our approach.

Crosslinguistic Influences Have Multiple Sources

The hierarchical inference framework naturally extends to the acquisition of L3 and beyond, predicting that any previously acquired language may affect learning of a new language.

Given that learners infer the underlying structure of their total linguistic environment, they must represent this information in a way that reflects the interconnectedness of the system. No language is a priori privileged as the source of transfer; rather, each previously acquired language contributes evidence toward the underlying structure of the environment. This does not mean that every language is expected to exert equal influence on the target L*n*, as the degree of influence will depend on other factors, such as between-language structural similarities (see below).

The hierarchical inference framework differs in this respect from other standard approaches to L2 acquisition, which do not have an obvious way of capturing the acquisition of L3 and beyond. When L1 properties are assumed to transfer to the L2 initial state at the onset of acquisition, it becomes unclear what is predicted in the case of a multilingual learner: Should transfer occur from L1, L2, or a combination of both? The most straightforward extension of these approaches would be to expect that L1 should be the main (or even only) source of transfer, just as in the case of L2 acquisition, but other interpretations are also possible (e.g., see Foote, 2009). Independent proposals have been developed in the field of third and additional language acquisition, investigating various factors that might determine the source of transfer, as discussed below. The main novel contribution of our framework is providing a principled way of deriving predictions for crosslinguistic influences in both L2 and L3/L*n* acquisition, in addition to unifying it with adaptation in L1.

The empirical findings regarding L3 acquisition are that transfer can apply from any previously learned language, whether native or nonnative (e.g., see de Bot & Jaensch, 2015; Rothman, Iverson, & Judy, 2011), which is precisely the prediction of the hierarchical inference framework. For example, beginner and intermediate learners of L3-Brazilian Portuguese with previous Spanish exposure utilize their knowledge of Spanish object clitic pronouns when learning similar clitic pronouns in Portuguese (whether Spanish is their L1 or L2), with English as L2 or L1, respectively (Montrul, Dias, & Santos, 2011). Another example comes from a large-scale study of over 50,000 learners of Dutch with varying language backgrounds, showing independent influence of both L1 and L2 on the attained proficiency in L3-Dutch (Schepens, Van der Slik, & Van Hout, submitted).

Crosslinguistic Influences Are Based on Perceived Similarities

In the hierarchical inference framework, the effect of previously learned languages depends on how close a given language is to the target language in the inferred similarity-based hierarchy, and how certain the learner is about a particular inferred relation between languages. Once a learner has observed some similarities between two languages, further similarities are hypothesized, because the learner has likely placed the two languages close to each other in the inferred hierarchy. This means that we expect to observe an overextension of properties from a known language to the target language as a function of the perceived similarity between languages, at least at the beginning of acquisition. As already discussed, the inferred similarity between languages depends on both the objective typological relationship and other factors that distort learners' perception of these similarities, such as learning two languages in similar contexts. Therefore, we predict more pervasive influence between languages that are typologically more similar, as well as those

that are alike in other respects, such as the environments in which they were learned (e.g., two nonnative languages). However, as learning progresses and learners uncover the properties of the new language, we expect actual typological similarities to play an increasingly prominent role, with other factors diminishing in their influence. Indeed, there is evidence that L2-to-L3 influence generally diminishes with increased L3 proficiency (e.g., Wrembel, 2010).

This aspect of the hierarchical inference framework is entirely consistent with the insights developed in a large body of research on L3 acquisition, investigating what factors—including between-language similarity—determine which of previously learned languages is the source of transfer to a new language (see Giancaspro, Halloran, & Iverson, 2015; Rothman, 2015). However, there are important differences between this previous work and our proposal. The hierarchical inference framework predicts that all previously learned languages affect transfer to a new language, and that each of these previously learned languages does so to the extent that learners implicitly perceive it to be similar to the new language. The previous work, on the other hand, has largely focused on determining a single most important factor in transfer. For example, some research has investigated whether the source language for transfer to a new language is always the typologically most similar language (e.g., Montrul et al., 2011; Rothman, 2011) or always another nonnative language (e.g., Bardel & Falk, 2007; Falk & Bardel, 2011).

The hierarchical inference framework may be able to reconcile these mixed findings and claims by providing a principled explanation of how different factors jointly contribute to the observed crosslinguistic influences. Additionally, the hierarchical inference framework predicts that the influence of a language will depend on the certainty that learners have in their indexical hierarchically structured implicit beliefs about this language, which is a function of the amount of previous exposure they have had to the language. This means that the shape of the inferred hierarchy is expected to change across L*n* acquisition. For example, at the early stages of L*n* acquisition, learners lack sufficient data from the target language to adequately assess its actual structural similarities to previously learned languages, and so they may overrely on other factors, such as presumed greater similarity between two nonnative languages (e.g., L2 and L3, due to similarities in the environments in which they were learned) than between the native and a nonnative language (e.g., L1 and L3). As learners receive more for input from the target language, they are expected to increasingly take into account the actual observed between-language similarities. Our proposal thus provides a testable guiding framework for future work on the relative influence of different previously learned languages in learning a new language. These predictions are shared with other accounts that emphasize the role of perceived between-language similarities or psychotypology (e.g., Rothman, 2015) but—in the hierarchical inference framework—they necessarily follow from the underlying architecture of hierarchical probabilistic inference.

The predictions of the hierarchical inference framework regarding similarity-based transfer are supported by existing findings. First, there is evidence that the benefit of L1 knowledge depends gradiently on the typological distance between L1 and L2 (Schepens, Van der Slik, & Van Hout, 2013). In particular, Schepens and colleagues examined the proficiency scores of over 50,000 learners with varying language backgrounds in an official state exam of

Dutch, and found that the scores covaried systematically with morphological similarities between Dutch and the learners' L1 (after controlling for other factors, such as length of residence in the Netherlands and age of arrival): The higher the between-language similarity, the higher the exam score. In addition, Schepens, Van der Slik, and Van Hout (2015, submitted) observed similar gradient effects of typological distance in the case of L3 acquisition when examining the L3-Dutch proficiency scores in relation to the similarities between Dutch and the learners' L2 (after controlling for other factors, including the learners' L1).

Second, the hierarchical inference framework naturally captures the rather surprising finding that learners sometimes fail to transfer the properties that are identical in one known language and the target language, and instead appear to transfer nontarget properties from another language—one that is, for instance, typologically closer. One example comes from the case of L1-English beginner learners of French in their use of subject pronouns (Rothman & Cabrelli Amaro, 2010). Both English and French are characterized by obligatory subject pronouns, and L1-English L2-French learners perform very well in their subject pronoun use in French. At the same time, equal-proficiency L3-French learners with previous knowledge of L2-Spanish frequently accept ungrammatical null-subject sentences in French. This result can be attributed to negative transfer from L2-Spanish, which is a language that allows subject pronoun dropping. Similar examples can be found for L1-Swedish L2-English L3-German learners in their verb placement (Bohnacker, 2006; Håkansson, Pienemann, & Sayehli, 2002). While both Swedish and German are verb-second languages, these learners produce fewer correct verb-second utterances in German than L1-Swedish L2-German learners with no prior exposure to English. Again, this can be attributed to the influence of L2-English, which—unlike other Germanic languages—is not characterized by the verb-second syntax. Within the hierarchical inference framework, this "transfer blocking by L2" (e.g., Bardel & Falk, 2007) is explained by learners' inferred close relationship between French and Spanish or German and English. There are multiple possible reasons why learners might be expected to infer such relationship in these cases: objective typological similarities, nonnative status of both languages, or perhaps even top-down beliefs that both languages belong to the same language group. Once learners establish that French and Spanish or German and English are close in the linguistic hierarchy, they overextend the similarities to the properties that are in fact different across the two languages.

Crosslinguistic Influences are Multidirectional

Another aspect of crosslinguistic influence expected within the hierarchical inference approach is its multidirectionality, where an L*n* can affect learners' previously acquired languages, including L1. This is because the learners' implicit beliefs capture the whole structure of their linguistic environment in a way that is interconnected. The interconnectedness is necessary because learners continuously adjust their inferences drawing on the total of their language knowledge. Therefore, it must be the case that inferences about L*n* should be able to affect previously learned languages in the same way that previously learned languages affect L*n*. The extent of this backward (or reverse) influence (e.g., L2-to-L1) depends on the same factors as the forward influence (e.g., L1-to-

L2): inferred between-language similarity as well as the degree of uncertainty about each model. It is noteworthy that well-established language representations (e.g., L1 or other languages with near-native proficiency) should be relatively more resistant to modifications than representations of languages about which learners have more uncertainty (e.g., low-proficiency L2 or attrited L1).

These predictions are consistent with the existing L2/L*n* acquisition data. First, there is evidence that a L3/L*n* can affect the learner's L2. For example, learning a L3 that allows null subjects influences the rate at which null subjects are accepted in the learner's L2. In particular, Aysan (2012) found that L1-Turkish L2-English learners accept more (ungrammatical) null-subject sentences in English when they also speak L3-Italian, which allows null subjects, relative to the case of no L3 or L3-French, which behaves like English in not allowing null subjects. Within the hierarchical inference framework, this can be explained by learners' strengthened beliefs about the optionality of subject pronouns in languages after having been exposed to Italian, which in turn leads to an adjustment of the previously learned grammar of English. Similarly, L1-Cantonese L2-English L3-German learners make mistakes in the tense/aspect use in English that can be traced back to the German grammar (e.g., using the present perfect tense for past events without current relevance), which is not observed for L1-Cantonese L2-English learners with no L3 or a non-Indo-European L3, such as Japanese, Korean, or Thai (Cheung, Matthews, & Tsang, 2011). The L3-to-L2 influence can also be beneficial. For example, showing an understanding of the perfective versus imperfective aspect distinction that exists in all Romance languages is superior in L1-English L2-Romance learners who also know another L3-Romance language (French, Italian, or Spanish) relative to L1-English L2-Romance learners with no L3 (Foote, 2009).

Second, the influence of nonnative languages extends even to the learner's L1. The extreme case of this influence is L1 attrition, which involves a simplification or an impairment of the L1 system, that is, inability to produce some L1 elements (e.g., Köpke, Schmid, Kejzer, & Dostert, 2007). Under this scenario, L_{any} inferences become gradually dominated by the learners' nonnative languages, leading to increasing adjustments to the L1 grammar, especially in cases when the dominant nonnative language is perceived as highly similar to the L1. However, small adjustments to L1 are also expected even when L1 is still used on a regular basis, and indeed researchers have identified other types of L2/L*n* influence that add to the L1 system without entailing the loss of the original L1 knowledge. Generally, the first signs of L*n* influence on L1 involve lexical borrowings, semantic extensions, and loan translation (see Pavlenko, 2000). For example, adult L1-Russian L2-English learners immersed in an English-speaking environment were found to use Russian words with broader semantic ranges that characterize their correspondent English equivalents (Pavlenko & Jarvis, 2002). L*n*-to-L1 influence has also been documented in other areas, including phonology, morphosyntax, conceptual representations, and pragmatics (e.g., Chang, 2012; Dmitrieva, Jongman, & Sereno, 2010; Mennen, 2004; Ulbrich & Ordin, 2014). For example, Dmitrieva et al. (2010) found that monolingual L1-Russian speakers use the duration of the release and closure/frication to distinguish voiceless and partially-devoiced word-final obstruents. However, adult L1-Russian L2-English learners immersed in an English-speaking environment use two additional cues that are also used in English to encode this

contrast. In a different domain, Tsimpli, Sorace, Heycock, and Filiaci (2004) demonstrated L2-to-L1 influence in L1-Italian and L1-Greek learners of L2-English immersed in an English-speaking environment for a minimum of six years, using both L1 and L2 on the daily basis. L1-Greek speakers were found to produce a higher rate of overt preverbal subjects in Greek than Greek monolinguals, and L1-Italian speakers inappropriately extended the scope of overt pronominal subjects in Italian, both of which can be attributed to the influence of English.

Statistical Knowledge Affects the Content of Crosslinguistic Influences

The final point concerns the exact content of transfer. While the hierarchical inference approach does not impose any a priori constraints in this regard, it is very much in line with recent findings suggesting that crosslinguistic transfer involves drawing not only on the specific categories that exist in the source language but also on the statistical distributions over those categories.

Some evidence for this comes from studies on the initial segmentation of words out of a continuous nonnative speech stream, showing that it is affected by the statistical regularities of the learners' L1. For example, during initial exposure to a new language, L1-Korean learners tend to rely on forward transitional probabilities between syllables, while L1-English learners tend to rely on backward probabilities (Onnis & Thiessen, 2013). This can be attributed to the fact that forward probabilities are generally more informative in Korean given its left-branching word order, while backward probabilities are more informative in English given its right-branching word order (see corpus analyses of both languages in Onnis & Thiessen, 2013). In a similar vein, L1-English learners segment words in a new language based on both transitional probabilities of the input and generalizations over L1 phonotactics (Finn & Hudson Kam, 2008); the influence of L1 phonotactics also extends to morphological learning (Finn & Hudson Kam, 2015). Finally, L1-Khalkha Mongolian learners are more sensitive to nonadjacent vocalic dependencies in a new language than L1-English or L1-French learners, which has been argued to arise from Khalkha vowel harmony patterns that are absent from English or French (LaCross, 2015). Similar results have also been observed in the domain of nonnative phonetic category learning, where the overall informativity of acoustic or articulatory cues in L1 affects the way those cues are weighed when processing and learning nonnative phonetic categories, either facilitating or hindering acquisition (e.g., Bohn & Best, 2012; Pajak & Levy, 2014).

All the above findings can be captured within the hierarchical inference framework, because learners are expected to draw on their prior beliefs in any way that provides them with the best possible guesses about the structure of the new language. This means that when interpreting the L*n* statistical properties, learners should be influenced not only by the specific categories that exist in the previously learned languages, but also by statistical distributions over those categories. This influence will lead to interference when, for example, the L2 statistical cues conflict with L1 properties (e.g., phonotactic constraints, phonetic categorization cues), because learners' expectations down-weight the statistical regularities found in the input. On the other hand, this bias can also lead to facilitation when the L2 statistical cues align with prior expectations. More generally, these biases allow

learners to take advantage of commonalities between languages—including, for example, those that stem from commonalities in the use of language. The original reason for the existence of such biases is, however, likely their necessity for robust L1 speech perception and processing (cf. Kleinschmidt & Jaeger, 2015).

Future Research

The hierarchical inference framework raises many new questions for future research. Here we briefly review three questions that we consider of particular interest. One question concerns the exact content and shape of L_{any} inferences. We view L_{any} as a distribution over language properties, encoding the information about the likelihood of different properties across languages. In particular, L_{any} inferences may consist of a range of linguistically relevant cues across different language domains (e.g., acoustic-phonetic features, word order, animacy, case inflection, etc.), where each cue is accompanied by a weight (or attention strength; cf. Bates & MacWhinney, 1987; Escudero & Boersma, 2004; MacWhinney, 1997, 2008). Within this L_{any} conceptualization, learners are expected to make inferences about possible languages that go beyond the properties of each individual language they know. However, the extent and nature of generalizations from prior linguistic beliefs is still not very well understood (see Pajak & Levy, 2014). The same problem arises within L1, for example when generalizing between speakers or dialects/accents (Kleinschmidt & Jaeger, 2015). Therefore, pinning down the nature of L_{any} inferences will only be possible by collecting more data pertinent to crosslinguistic generalization patterns.

Another open question of great theoretical relevance concerns the way in which learners capture the hierarchical statistical structure of their linguistic environment. One possibility is that it is based on the overall similarity between languages (i.e., learners adopt the assumption that all features are either similar or not between languages), as we proposed here. The main reason to expect that this may be the right approach is that it is a simplifying assumption that allows learners to pool all their data, thus leading to more confident (though less accurate) estimates of similarity across features. This may be especially useful at the early stages of L*n* acquisition, when evidence from L*n* input is highly limited. However, it may be that learners capture the hierarchical statistical structure relative to a linguistic category: for example, that L1 and L2 are similar with regard to how they realize voicing, but differ with regard to how they encode grammatical function assignment. Yet aiming to capture the hierarchical statistics of every cue would quickly lead to data sparseness, which might not allow learners to make any potentially useful generalizations. The two possibilities outlined above are not necessarily incompatible. In fact, it is likely that the way learners capture the statistical structure of their environment changes across L*n* acquisition. For example, learners might begin L*n* acquisition with a simplified measure of overall similarities between languages, which allows them to make quick generalizations at the onset of learning. Later during acquisition, however, when learners already have access to a larger amount of evidence about the target L*n*, they may transition to a more refined encoding of similarities that is based on individual linguistic categories. This would let multilingual learners take advantage of similarities between different sets of languages for each specific aspect of the language they try to acquire (see Rothman, 2015).

Finally, in this article we largely focused on between-language transfer during learning. However, the way learners capture the structure of their linguistic environment is likely to also affect their inferences during online language production and comprehension. In fact, it might be more intuitive to think of some aspects of transfer as happening purely during processing due to languages coexisting in the brain and being coactivated (for a review, see Kroll, Bobb, & Hoshino, 2014), as evinced, for example, in lexical intrusions (e.g., Poulisse & Bongaerts, 1994) or sound productions that appear to be a mixture of two languages (e.g., Wunder, 2011). Other processes, on the other hand, may be more intuitively interpreted as changes to the mental representations of each language, their mutual strengths, the relations between them, or how these representations are accessed (e.g., Amaral & Roeper, 2014). A good case in point, for example, would be facilitation in understanding the perfective versus imperfective aspect distinction in L3-Italian due to the knowledge of L2-Spanish (Foote, 2009). In our view, both of these two types of crosslinguistic influence play a role, and investigating how they interact is an important area for future work.

Conclusion

We presented a new hierarchical inference framework to investigate the role of prior language knowledge in L2/L*n* acquisition. The framework has two crucial components: (a) statistical learning as one of the mechanisms through which adults acquire new languages, and (b) representations of language knowledge that captures the hierarchically structured linguistic environment of bi/multilingual learners. We proposed that in addition to the representations of each acquired language, learners also make higher-level inferences about what linguistic structures are likely in any language. We further proposed that learning proceeds through probabilistic inference under uncertainty. That is, learners combine new language input with their prior language knowledge, and make inferences about the underlying structure of the language they are learning, while at the same time adjusting their beliefs about any language. We motivated this framework in recent research on L1 perception and sentence understanding, and argued that the same architecture—hierarchically organized language models—captures both L1 and L2/L*n* processing and learning. Our proposal builds on a large body of prior work in different domains, bringing together insights that, as we argued, are of great relevance to L2/L*n* research. The hierarchical inference framework (a) provides a unified view of both L1 adaptation and L2/L*n* learning as continuous probabilistic inferences in response to language input, and (b) helps reconceptualize the nature of transfer in L2/L*n* acquisition by viewing it as learners' inferences about the target language based on their current total language knowledge. In this way, our approach extends previous proposals, such as Ellis' emergentist account (Ellis, 2006a, b; Ellis, O'Donnel, & Römer, 2013) or MacWhinney's Unified Model (MacWhinney, 2008, 2012).

Acknowledgments

The authors would like to thank four anonymous reviewers, whose comments and suggestions have been extremely helpful in revising the manuscript. We are especially grateful to Lourdes Ortega, who went far beyond her duty in providing invaluable insights and assisting us with the revisions. This research was supported by an NIH post-doctoral fellowship to BP (NIH Training Grant T32-DC000035 awarded to the Center for Language Sciences at University of Rochester), an NSF Graduate Research Fellowship to DK, an NIH post-doctoral fellowship to AF (NIH Training Grant T32-HD055272), and by the Eunice Kennedy Shriver National Institute of Child Health &

Human Development of the National Institutes of Health under Award Number R01HD075797 to TFJ. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Abrahamsson N, Hyltenstam K. The robustness of aptitude effects in near-native second language acquisition. Studies in Second Language Acquisition. 2008; 30:481–509. doi:10.1017/S027226310808073X.

Allen JS, Miller JL, DeSteno D. Individual talker differences in voice-onset-time. Journal of the Acoustical Society of America. 2003; 113:544–552. doi:10.1121/1.1528172. [PubMed: 12558290]

Amaral L, Roeper T. Multiple grammars and second language representation. Second Language Research. 2014; 30:3–36. doi:10.1177/0267658313519017.

Anderson, JR. The adaptive character of thought. Lawrence Erlbaum; Hillsdale, NJ: 1990.

Arai M, Keller F. The use of verb-specific information for prediction in sentence processing. Language and Cognitive Processes. 2013; 28:525–560. doi:10.1080/01690965.2012.658072.

Aysan, Z. Reverse interlanguage transfer: The effects of L3 Italian & L3 French on L2 English pronoun use. Bilkent University; Turkey: 2012. Unpublished master's thesis

Baese-Berk MM, Bradlow AR, Wright BA. Accent-independent adaptation to foreign accented speech. JASA Express Letters. 2013; 133:174–180. doi:10.1121/1.4789864.

Bardel C, Falk Y. The role of the second language in third language acquisition: The case of Germanic syntax. Second Language Research. 2007; 24:459–484. doi:10.1177/0267658307080557.

Bates, E., MacWhinney, B. Competition, variation, and language learning.. In: MacWhinney, B., editor. Mechanisms of language acquisition. Lawrence Erlbaum; Hillsdale, NJ: 1987. p. 157-194.

Bejjanki VR, Clayards M, Knill DC, Aslin RN. Cue integration in categorical tasks: Insights from audio-visual speech perception. PLoS ONE. 2011; 6:e19812. doi:10.1371/journal.pone.0019812. [PubMed: 21637344]

Bertelson P, Vroomen J, de Gelder B. Visual recalibration of auditory speech identification: a McGurk aftereffect. Psychological Science. 2003; 14:592–597. doi:10.1046/j.0956-7976.2003.psci_1470.x. [PubMed: 14629691]

Best, CT., Shaw, JA., Docherty, G., Evans, BG., Foulkes, P., Hay, J., Al-Tamimi, J., Mair, K., Mulak, KE., Wood, S. From Newcastle MOUTH to Aussie ears: Australians' perceptual assimilation and adaptation for Newcastle UK vowels.. Proceedings of Interspeech; Dresden, Germany. 2015; 2015.

Birdsong, D. Age and the end state of second language acquisition.. In: Ritchie, WC., Bhatia, TK., editors. The new handbook of second language acquisition. Emerald; Bingley: 2009. p. 401-424.

Bohn OS, Best CT. Native-language phonetic and phonological influences on perception of American English approximants by Danish and German listeners. Journal of Phonetics. 2012; 40:109–128. doi:10.1016/j.wocn.2011.08.002.

Bohnacker U. When Swedes begin to learn German: From V2 to V2. Second Language Research. 2006; 22:443–486. doi:10.1191/0267658306sr275oa.

Boston MF, Hale J, Kliegl R, Patil U, Vasishth S. Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. Journal of Eye Movement Research. 2008; 2:1–12.

Bradlow AR, Bent T. Perceptual adaptation to non-native speech. Cognition. 2008; 106:707–729. doi:10.1016/j.cognition.2007.04.005. [PubMed: 17532315]

Bradlow AR, Akahane-Yamada R, Pisoni DB, Tohkura Y. Training Japanese listeners to identify English /r/and /l/: Long-term retention of learning in perception and production. Perception and Psychophysics. 1999; 61:977–985. doi:10.3758/BF03206911. [PubMed: 10499009]

Chang CB. Rapid and multifaceted effects of second-language learning on first-language speech production. Journal of Phonetics. 2012; 40:249–268. doi:10.1016/j.wocn.2011.10.007.

Cheung, ASC., Matthews, S., Tsang, WL. Transfer from L3 German to L2 English in the domain of tense/aspect.. In: De Angelis, G., Dewaele, J-M., editors. Second language acquisition: New trends in crosslinguistic influence and multilingualism research. Channel View Publications; Bristol, UK: 2011. p. 53-73.

Clayards MA, Tanenhaus MK, Aslin RN, Jacobs RA. Perception of speech reflects optimal use of probabilistic speech cues. Cognition. 2008; 108:804–809. doi:10.1016/j.cognition.2008.04.004. [PubMed: 18582855]

Creel SC, Aslin RN, Tanenhaus MK. Heeding the voice of experience: The role of talker variation in lexical access. Cognition. 2008; 106:633–664. doi:10.1016/j.cognition.2007.03.013. [PubMed: 17507006]

Culbertson J, Smolensky P, Legendre G. Learning biases predict a word order universal. Cognition. 2012; 122:306–329. doi:10.1016/j.cognition.2011.10.017. [PubMed: 22208785]

Cutler, A., Eisner, F., McQueen, JM., Norris, D. How abstract phonemic categories are necessary for coping with speaker-related variation.. In: Fougeron, C.Kühnert, B.D'Imperio, M., Vallée, N., editors. Laboratory phonology 10. De Gruyter Mouton; Berlin: 2010. p. 91-111.

Dahan D, Magnuson JS, Tanenhaus MK. Time course of frequency effects in spoken-word recognition: Evidence from eye movements. Cognitive Psychology. 2001; 42:317–367. doi:10.1006/cogp.2001.0750. [PubMed: 11368527]

De Angelis G. Interlanguage transfer of function words. Language Learning. 2005; 55:379–414. doi:10.1111/j.0023-8333.2005.00310.x.

de Bot K, Jaensch C. What is special about L3 processing? Bilingualism: Language and Cognition. 2015; 18:130–144. doi:10.1017/S1366728913000448.

Demberg V, Keller F. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. Cognition. 2008; 109:193–210. doi:10.1016/j.cognition.2008.07.008. [PubMed: 18930455]

Dikker S, Pylkkänen L. Predicting language: MEG evidence for lexical preactivation. Brain and Language. 2013; 127:55–64. doi:10.1016/j.bandl.2012.08.004. [PubMed: 23040469]

Dmitrieva O, Jongman A, Sereno J. Phonological neutralization by native and non-native speakers: The case of Russian final devoicing. Journal of Phonetics. 2010; 38:483–492. doi:10.1016/j.wocn.2010.06.001.

Eisner F, McQueen JM. Perceptual learning in speech: Stability over time. Journal of the Acoustical Society of America. 2006; 119:1950–1953. doi:10.1121/1.2178721. [PubMed: 16642808]

Ellis NC. Language acquisition as rational contingency learning. Applied Linguistics. 2006a; 27:1–24. doi:10.1093/applin/ami038.

Ellis NC. Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. Applied Linguistics. 2006b; 27:164–194. doi:10.1093/applin/aml015.

Ellis NC, O'Donnel MB, Römer U. Usage-based language: Investigating the latent structures that underpin acquisition. Language Learning. 2013; 63:25–51. doi:10.1111/j.1467-9922.2012.00736.x.

Endress AD, Mehler J. The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. Journal of Memory and Language. 2009; 60:351–367. doi:10.1016/j.jml.2008.10.003.

Escudero P, Boersma P. Bridging the gap between L2 speech perception research and phonological theory. Studies in Second Language Acquisition. 2004; 26:551–585. doi:10+10170S0272263104040021.

Escudero P, Williams D. Distributional learning has immediate and long-lasting effects. Cognition. 2014; 133:408–413. doi:10.1016/j.cognition.2014.07.002. [PubMed: 25128798]

Escudero P, Benders T, Wanrooij K. Enhanced bimodal distributions facilitate the learning of second language vowels. Journal of the Acoustical Society of America. 2011; 130:EL206–EL212. doi:10.1121/1.3629144. [PubMed: 21974493]

Falk Y, Bardel C. Stable and developmental optionality in native and non-native Hungarian grammars. Second Language Research. 2011; 27:59–82. doi:10.1177/0267658310386647.

Farmer, TA., Fine, AB., Yan, S., Cheimariou, S., Jaeger, TF. Error-driven adaptation of higher-level expectations during reading.. In: Bello, P.Guarini, M.McShane, M., Scassellati, B., editors. Proceedings of the 36th Annual Meeting of the Cognitive Science Society. Cognitive Science Society; Austin, TX: 2014. p. 2181-2186.

Farmer TA, Monaghan P, Misyak JB, Christiansen MH. Phonological typicality influences sentence processing in predictive contexts: A reply to Staub et al. (2009). Journal of Experimental Psychology: Learning, Memory, and Cognition. 2011; 37:1318–1325. doi:10.1037/a0023063.

Fedzechkina M, Jaeger TF, Newport EL. Language learners restructure their input to facilitate efficient communication. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109:17897–17902. doi:10.1073/pnas.1215776109. [PubMed: 23071337]

Feldman NH, Griffiths TL, Morgan JL. The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. Psychological Review. 2009; 116:752–782. doi:10.1037/a0017196. [PubMed: 19839683]

Fine AB, Jaeger TF, Farmer TA, Qian T. Rapid expectation adaptation during syntactic comprehension. PLoS ONE. 2013; 8:1–18. doi:10.1371/journal.pone.0077661.

Fine, AB., Qian, T., Jaeger, TF., Jacobs, RA. Proceedings of the 1st ACL Workshop on Cognitive Modeling and Computational Linguistics. Association for Computational Linguistics; 2010. Syntactic adaptation in language comprehension.; p. 18-26.

Finn AS, Hudson Kam CL. The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. Cognition. 2008; 108:477–499. doi:10.1016/j.cognition.2008.04.002. [PubMed: 18533142]

Finn AS, Hudson Kam CL. Why segmentation matters: Experience-driven segmentation errors impair “morpheme” learning. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2015; 41:1560–1569. doi:10.1037/xlm0000114.

Flege, JE. Age of learning and second-language speech.. In: Birdsong, DP., editor. Second language acquisition and the critical period hypothesis. Lawrence Erlbaum; Hillsdale, NJ: 1999. p. 101-132.

Foote, R. Transfer in L3 acquisition: The role of typology.. In: Leung, YI., editor. Third language acquisition and Universal Grammar. Multilingual Matters; Bristol, UK: 2009. p. 89-114.

Gatbonton E, Trofimovich P, Magid M. Learners' ethnic group affiliation and L2 pronunciation accuracy: A sociolinguistic investigation. TESOL Quarterly. 2005; 39:489–511. doi:10.2307/3588491.

Gebhart AL, Aslin RN, Newport E. Changing structures in midstream: Learning along the statistical garden path. Cognitive Science. 2009; 33:1087–1116. doi:10.1111/j.1551-6709.2009.01041.x. [PubMed: 20574548]

Giancaspro D, Halloran B, Iverson M. Transfer at the initial stages of L3 Brazilian Portuguese: A look at three groups of English/Spanish bilinguals. Bilingualism: Language and Cognition. 2015; 18:191–207. doi:10.1017/S1366728914000339.

Goldinger SD. Words and voices: Episodic traces in spoken word identification and recognition memory. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1996; 22:1166–1183. doi:10.1037/0278-7393.22.5.1166.

Goldinger SD. Echoes of echoes? An episodic theory of lexical access. Psychological Review. 1998; 105:251–279. doi:10.1037/0033-295X.105.2.251. [PubMed: 9577239]

Goudbeek M, Cutler A, Smits R. Supervised and unsupervised learning of multidimensionally varying non-native speech categories. Speech Communication. 2008; 50:109–125. doi:10.1016/j.specom.2007.07.003.

Griffiths TL, Vul E, Sanborn AN. Bridging levels of analysis for probabilistic models of cognition. Current Directions in Psychological Science. 2012; 21:263–268. doi:10.1177/0963721412447619.

Grodner, D., Sedivy, J. The effect of speaker-specific information on pragmatic inferences.. In: Gibson, E., Pearlmutter, N., editors. The processing and acquisition of reference. Vol. 2327. MIT Press; Cambridge, MA: 2011. p. 239-272.

Håkansson G, Pienemann M, Sayehli S. Transfer and typological proximity in the context of second language processing. Second Language Research. 2002; 18:250–273. doi:10.1191/0267658302sr206oa.

Hakuta K, Bialystok E, Wiley E. Critical evidence: A test of the critical-period hypothesis for second-language acquisition. Psychological Science. 2003; 14:31–38. doi:10.1111/1467-9280.01415. [PubMed: 12564751]

Han, Z-H. Fossilization in adult second language acquisition. Multilingual Matters; Clevedon, UK: 2004.

Hanulikova A, Van Alphen PM, Van Goch M, Weber A. When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. Journal of Cognitive Neuroscience. 2012; 24:878–887. doi:10.1162/jocn_a_00103. [PubMed: 21812565]

Hudson Kam CL. More than words: Adults learn probabilities over categories and relationships between them. Language Learning and Development. 2009; 5:115–145. doi: 10.1080/15475440902739962. [PubMed: 20161375]

Idemaru K, Holt LL. Word recognition reflects dimension-based statistical learning. Journal of Experimental Psychology: Human Perception and Performance. 2011; 37:1939–1956. doi: 10.1037/a0025641. [PubMed: 22004192]

Johnson JS, Newport EL. Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. Cognitive Psychology. 1989; 21:60–99. doi:10.1016/0010-0285(89)90003-0. [PubMed: 2920538]

Johnson, K. Speech perception without speaker normalization: An exemplar model.. In: Johnson, K., Mullennix, JW., editors. Talker variability in speech processing. Academic Press; San Diego, CA: 1997. p. 145-165.

Johnson, K. Speaker normalization in speech perception.. In: Pisoni, DB., Remez, RE., editors. The Handbook of speech perception. Blackwell Publishers; Oxford: 2005. p. 363-389.

Johnson, K. Factors that affect phonetic adaptation: Exemplar filters and sound change.. Talk presented at the Workshop on Current Issues and Methods in Speaker Adaptation; Columbus, OH.. 2013.

Johnson K, Strand E, D'Imperio M. Auditory-visual integration of talker gender in vowel perception. Journal of Phonetics. 1999; 27:359–384. doi:10.1006/jpho.1999.0100.

Kamide Y. Learning individual talkers' structural preferences. Cognition. 2012; 124:66–71. doi: 10.1016/j.cognition.2012.03.001. [PubMed: 22498776]

Kleinschmidt, DF., Fine, AB., Jaeger, TF. A belief-updating model of adaptation and cue combination in syntactic comprehension.. In: Miyake, N.Peebles, D., Cooper, RP., editors. Proceedings of the 34th Annual Conference of the Cognitive Science Society. Cognitive Science Society; Austin, TX: 2012. p. 599-604.

Kleinschmidt, D., Jaeger, TF. Proceedings of the 2nd ACL Workshop on Cognitive Modeling and Computational Linguistics. Association for Computational Linguistics; Stroudsburg, PA: 2011. A Bayesian belief updating model of phonetic recalibration and selective adaptation..

Kleinschmidt, DF., Jaeger, TF. A continuum of phonetic adaptation: Evaluating an incremental belief-updating model of recalibration and selective adaptation.. In: Miyake, N.Peebles, D., Cooper, RP., editors. Proceedings of the 34th Annual Conference of the Cognitive Science Society. Cognitive Science Society; Austin, TX: 2012. p. 605-610.

Kleinschmidt DF, Jaeger TF. Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. Psychological Review. 2015; 122:148–203. doi:10.1037/a0038695. [PubMed: 25844873]

Kondaurova MV, Francis AL. The role of selective attention in the acquisition of English tense and lax vowels by native Spanish listeners: Comparison of three training methods. Journal of Phonetics. 2010; 38:569–587. doi:10.1016/j.wocn.2010.08.003. [PubMed: 21499531]

Köpke, B.Schmid, MS.Kejzer, M., Dostert, S., editors. Language attrition: Theoretical perspectives. John Benjamins; Philadelphia, PA: 2007.

Kraljic T, Samuel AG. Perceptual learning for speech: Is there a return to normal? Cognitive Psychology. 2005; 51:141–178. doi:10.1016/j.cogpsych.2005.05.001. [PubMed: 16095588]

Kraljic T, Samuel AG. Generalization in perceptual learning for speech. Psychonomic Bulletin & Review. 2006; 13:262–268. doi:10.3758/BF03193841. [PubMed: 16892992]

Kraljic T, Brennan SE, Samuel AG. Accommodating variation: Dialects, idiolects, and speech processing. Cognition. 2008; 107:51–81. doi:10.1016/j.cognition.2007.07.013.

Kroll JF, Bobb SC, Hoshino N. Two languages in mind: Bilingualism as a tool to investigate language, cognition, and the brain. Current Directions in Psychological Science. 2014; 23:159–163. doiI: 10.1177/0963721414528511. [PubMed: 25309055]

Kuperberg G, Jaeger TF. What do we mean by prediction in language comprehension? Language, Cognition, and Neuroscience. 2015 Published online 13 November 2015. doi: 10.1080/23273798.2015.1102299.

Kurumada, C. Contextual inferences over speakers' pragmatic intentions: Preschoolers' comprehension of contrastive prosody.. In: Knauff, M.Pauen, M.Sebanz, N., Wachsmuth, I., editors. Proceedings of the 35th Annual Conference of the Cognitive Science Society. Cognitive Science Society; Austin, TX: 2013. p. 852-857.

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D., Tanenhaus, MK. Rapid adaptation in online pragmatic interpretation of contrastive prosody.. In: Bello, P.Guarini, M.McShane, M., Scassellati, B., editors. Proceedings of the 36th Annual Meeting of the Cognitive Science Society. Cognitive Science Society; Austin, TX: 2014. p. 791-796.

LaCross A. Khalkha Mongolian speakers' vowel bias: L1 influences on the acquisition of non-adjacent vocalic dependencies. Language, Cognition, and Neuroscience. 2015; 30:1033–1047. doi: 10.1080/23273798.2014.915976.

Lewis R, Howes A, Singh S. Computational rationality: Linking mechanism and behavior through bounded utility maximization. Topics in Cognitive Science. 2014; 6:279–311. doi:10.1111/tops. 12086. [PubMed: 24648415]

Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. Psychological Review. 1967; 74:431–461. doi:10.1037/h0020279. [PubMed: 4170865]

Lim S-J, Holt L. Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization. Cognitive Science. 2011; 35:1390–1405. doi:10.1111/j. 1551-6709.2011.01192.x. [PubMed: 21827533]

Luce PA, Pisoni DB. Recognizing spoken words: The neighborhood activation model. Ear and Hearing. 1998; 19:1–36. doi:10.1097/00003446-199802000-00001. [PubMed: 9504270]

MacDonald MC. How language production shapes language form and comprehension. Frontiers in Psychology. 2013; 4:1–16. doi:10.3389/fpsyg.2013.00226. [PubMed: 23382719]

MacDonald MC, Just MA, Carpenter PA. Working memory constraints on the processing of syntactic ambiguity. Cognitive Psychology. 1992; 24:56–98. doi:10.1016/0010-0285(92)90003-K. [PubMed: 1537232]

MacDonald MC, Pearlmutter N, Seidenberg MS. The lexical nature of syntactic ambiguity resolution. Psychological Review. 1994; 101:676–703. doi:10.1037/0033-295X.101.4.676. [PubMed: 7984711]

MacWhinney B. Miniature linguistic systems as tests of the use of universal operating principles in second-language learning by children and adults. Journal of Psycholinguistic Research. 1983; 12:467–478. doi:10.1007/BF01068027.

MacWhinney, B. Second language acquisition and the Competition Model.. In: De Groot, AMB., Kroll, JF., editors. Tutorials in bilingualism: Psycholinguistic perspectives. Lawrence Erlbaum; Mahwah, NJ: 1997. p. 113-142.

MacWhinney, B. A unified model.. In: Robinson, P., Ellis, N., editors. Handbook of cognitive linguistics and second language acquisition. Lawrence Erlbaum; Mahwah, NJ: 2008. p. 341-371.

MacWhinney, B. The logic of the Unified Model.. In: Gass, SM., Mackey, A., editors. Handbook of second language acquisition. Routledge; New York: 2012. p. 211-227.

Marinova-Todd SH, Marshall DB, Snow CE. Three misconceptions about age and L2 learning. TESOL Quarterly. 2000; 34:9–34.

Maye J, Aslin RN, Tanenhaus MK. The weckud wetch of the wast: Lexical adaptation to a novel accent. Cognitive Science. 2008; 32:543–562. doi:10.1080/03640210802035357. [PubMed: 21635345]

McClelland JL, Elman JL. The TRACE model of speech perception. Cognitive Psychology. 1986; 18:1–86. doi:10.1016/0010-0285(86)90015-0. [PubMed: 3753912]

McClelland, JL., Thomas, A., McCandliss, BD., Fiez, JA. Understanding failures of learning: Hebbian learning, competition for representational space, and some preliminary experimental data.. In: Reggia, J.Ruppin, E., Glanzman, D., editors. Brain, behavioral, and cognitive disorders: The neurocomputational perspective. Elsevier; Oxford: 1999. p. 75-80.

McDonald SA, Shillcock RC. Eye movements reveal the on-line computation of lexical probabilities during reading. Psychological Science. 2003; 14:648–652. doi:10.1046/j. 0956-7976.2003.psci_1480.x. [PubMed: 14629701]

McMurray B, Jongman A. What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. Psychological Review. 2011; 118:219–246. doi:10.1037/a0022325. [PubMed: 21417542]

McQueen JM, Cutler A, Norris D. Phonological abstraction in the mental lexicon. Cognitive Science. 2006; 30:1113–1126. doi:10.1207/s15516709cog0000_79. [PubMed: 21702849]

Mennen I. Bi-directional interference in the intonation of Dutch speakers of Greek. Journal of Phonetics. 2004; 32:543–563. doi:10.1016/j.wocn.2004.02.002.

Metzing C, Brennan SE. When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. Journal of Memory and Language. 2003; 49:201–213. doi:10.1016/S0749-596X(03)00028-7.

Miyawaki K, Strange W, Verbrugge RR, Liberman AM, Jenkins JJ, Fujimura O. An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. Perception and Psychophysics. 1975; 18:331–340. doi:10.3758/BF03211209.

Montrul S, Dias R, Santos H. Clitics and object expression in the L3 acquisition of Brazilian Portuguese: Structural similarity matters for transfer. Second Language Research. 2011; 27:21–58. doi:10.1177/0267658310386649.

Moyer A. Do language attitudes determine accent? A study of bilinguals in the USA. Journal of Multilingual and Multicultural Development. 2007; 28:502–518. doi:10.2167/jmmd514.0.

Moyer A. Exceptional outcomes in L2 phonology: The critical factors of learner engagement and self-regulation. Applied Linguistics. 2014; 35:418–440. doi:10.1093/applin/amu012.

Myslín M, Levy R. Comprehension priming as rational expectation for repetition: Evidence from syntactic processing. Cognition. 2016; 147:29–56. doi:10.1016/j.cognition.2015.10.021. [PubMed: 26605963]

Nearey TM. The segment as a unit of speech perception. Journal of Phonetics. 1990; 18:347–373.

Nearey TM. Speech perception as pattern recognition. Journal of the Acoustical Society of America. 1997; 101:3241–3254. doi:10.1121/1.418290. [PubMed: 9193041]

Nearey TM, Assmann PF. Modeling the role of inherent spectral change in vowel identification. Journal of the Acoustical Society of America. 1986; 80:1297–1308. doi:10.1121/1.394433.

Nearey, TM., Hogan, JT. Phonological contrast in experimental phonetics: Relating distributions of production data to perceptual categorization curves.. In: Ohala, JJ., Jaeger, JJ., editors. Experimental phonology. Academic Press; Orlando, FL: 1986. p. 141-161.

Newman RS, Clouse S, Burnham JL. The perceptual consequences of within-talker variability in fricative production. Journal of the Acoustical Society of America. 2001; 109:1181–1196. doi: 10.1121/1.1348009. [PubMed: 11303932]

Newport EL, Aslin RN. Learning at a distance I: Statistical learning of non-adjacent dependencies. Cognitive Psychology. 2004; 48:127–162. doi:10.1016/S0010-0285(03)00128-2. [PubMed: 14732409]

Niedzielski N. The effect of social information on the perception of sociolinguistic variables. Journal of Language and Social Psychology. 1999; 18:62–85. doi:10.1177/0261927X99018001005.

Nielsen, K., Wilson, C. A hierarchical Bayesian model of multi-level phonetic imitation.. In: Abner, N., Bishop, J., editors. Proceedings of the 27th West Coast Conference on Formal Linguistics. Cascadilla Proceedings Project; Somerville, MA: 2008. p. 335-343.

Norris D, McQueen JM. Shortlist B: A Bayesian model of continuous speech recognition. Psychological Review. 2008; 115:357–395. doi:10.1037/0033-295X.115.2.357. [PubMed: 18426294]

Norris D, McQueen JM, Cutler A. Perceptual learning in speech. Cognitive Psychology. 2003; 47:204–238. doi:10.1016/S0010-0285(03)00006-9. [PubMed: 12948518]

O'Grady W. The emergentist program. Lingua. 2008; 118:447–464. doi:10.1016/j.lingua.2006.12.001.

Odlin, T. Crosslinguistic influence in second language acquisition.. In: Chapelle, CA., editor. The encyclopedia of applied linguistics. Blackwell; Malden, MA: 2013. p. 1562-1568.

Onishi KH, Chambers KE, Fisher C. Learning phonotactic constraints from brief auditory experience. Cognition. 2002; 83:B13–B23. doi:10.1016/S0010-0277(01)00165-2. [PubMed: 11814489]

Onnis L, Thiessen E. Language experience changes subsequent learning. Cognition. 2013; 126:268–284. doi:10.1016/j.cognition.2012.10.008. [PubMed: 23200510]

Pajak, B. Inductive inference in non-native speech processing and learning. University of California; San Diego: 2012. Unpublished doctoral dissertation

Pajak, B., Levy, R. Phonological generalization from distributional evidence.. In: Carlson, L.Hölscher, C., Shipley, T., editors. Proceedings of the 33rd Annual Conference of the Cognitive Science Society. Cognitive Science Society; Austin, TX: 2011. p. 2673-2678.

Pajak B, Levy R. The role of abstraction in non-native speech perception. Journal of Phonetics. 2014; 46:147–160. doi:10.1016/j.wocn.2014.07.001. [PubMed: 25197153]

Papp S. Stable and developmental optionality in native and non-native Hungarian grammars. Second Language Research. 2000; 16:173–200. doi:10.1191/026765800666966395.

Pavlenko A. L2 influence on L1 in late bilingualism. Issues in Applied Linguistics. 2000; 11:175–205.

Pavlenko A, Jarvis S. Bidirectional transfer. Applied Linguistics. 2002; 23:190–214. doi:10.1093/applin/23.2.190.

Peterson GE, Barney HL. Control methods used in a study of the vowels. Journal of the Acoustical Society of America. 1952; 24:175–184. doi:10.1121/1.1906875.

Pierrehumbert JB. Phonetic diversity, statistical learning, and acquisition of phonology. Language and Speech. 2003; 46:115–154. doi:10.1177/00238309030460020501. [PubMed: 14748442]

Poulisse N, Bongaerts T. First language use in second language production. Applied Linguistics. 1994; 15:36–57. doi:10.1093/applin/15.1.36.

Qian T, Jaeger TF, Aslin RN. Learning to represent a multi-context environment: More than detecting changes. Frontiers in Psychology. 2012; 3:228. doi:10.3389/fpsyg.2012.00228. [PubMed: 22833727]

Rebuschat, P. Implicit and explicit learning of languages. John Benjamins; Amsterdam: 2015.

Reeder PA, Newport EL, Aslin RN. From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. Cognitive Psychology. 2013; 66:30–54. doi:10.1016/j.cogpsych.2012.09.001. [PubMed: 23089290]

Rothman J. L3 syntactic transfer selectivity and typological determinacy: The typological primacy model. Second Language Research. 2011; 27:107–127. doi:10.1177/0267658310386439.

Rothman J. Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: Timing of acquisition and proficiency considered. Bilingualism: Language and Cognition. 2015; 18:179–190. doi:10.1017/S136672891300059X.

Rothman J, Cabrelli Amaro J. What variables condition syntactic transfer? A look at the L3 initial state. Second Language Research. 2010; 26:189–218. doi:10.1177/0267658309349410.

Rothman J, Iverson M, Judy T. Introduction: Some notes on the generative study of L3 acquisition. Second Language Research. 2011; 27:5–19. doi:10.1177/0267658310386443.

Saffran JR, Newport EL, Aslin RN. Word segmentation: The role of distributional cues. Journal of Memory and Language. 1996; 35:606–621. doi:10.1006/jmla.1996.0032.

Schepens J, Van der Slik F, Van Hout R. Learning complex features: Morphological account of L2 learnability. Language Dynamics and Change. 2013; 3:218–244. doi:10.1163/22105832-13030203.

Schepens J, Van der Slik F, Van Hout R. L1 and L2 distance effects in learning L3 Dutch. Language Learning. 2015 Published online 8 October 2015. doi:10.1111/lang.12150.

Schepens, J., Van der Slik, F., Van Hout, R. The L2 impact on acquiring Dutch as an L3: The L2 distance effect.. In: Speelman, D.Heylen, K., Geeraerts, D., editors. Mixed effects regression models in linguistics. Springer; Berlin: submitted

Selinker L. Interlanguage. International Review of Applied Linguistics. 1972; 10:209–231. doi:10.1515/iral.1972.10.1-4.209.

Selinker, L. Rediscovering interlanguage. Longman; London: 1992.

Sharwood Smith, M., Truscott, J. The multilingual mind: A modular processing perspective. Cambridge University Press; New York: 2014.

Smith NJ, Levy R. The effect of word predictability on reading time is logarithmic. Cognition. 2013; 128:302–319. doi:10.1016/j.cognition.2013.02.013. [PubMed: 23747651]

Smits R. Evidence for hierarchical categorization of coarticulated phonemes. Journal of Experimental Psychology: Human Perception and Performance. 2001a; 27:1145–1162. doi: 10.1037/0096-1523.27.5.1145. [PubMed: 11642700]

Smits R. Hierarchical categorization of coarticulated phonemes: A theoretical analysis. Perception & Psychophysics. 2001b; 63:1109–1139. doi:10.3758/BF03194529. [PubMed: 11766939]

Staum Casasanto, L. Does social information influence sentence processing?. In: Love, BC.McRae, K., Sloutsky, VM., editors. Proceedings of the 30th Annual Meeting of the Cognitive Science Society. Cognitive Science Society; Austin, TX: 2008. p. 799-804.

Stevens G. Age at immigration and second language proficiency among foreignborn adults. Language in Society. 1999; 28:555–578.

Strand EA. Uncovering the role of gender stereotypes in speech perception. Journal of Language and Social Psychology. 1999; 18:86–100. doi:10.1177/0261927X99018001006.

Tabor W, Juliano CJ, Tanenhaus MK. Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. Language and Cognitive Processes. 1997; 12:211–271. doi:10.1080/016909697386853.

Tanenhaus MK, Spivey-Knowlton MJ, Eberhard K, Sedivy J. Integration of visual and linguistic information in spoken language comprehension. Science. 1995; 268:1632–1634. doi:10.1126/science.7777863. [PubMed: 7777863]

Trueswell JC, Tanenhaus MK, Kello C. Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. Journal of Experimental Psychology: Learning, Memory and Cognition. 1993; 19:528–553. doi:10.1037/0278-7393.19.3.528.

Tsimpli I, Sorace A, Heycock C, Filiaci F. First language attrition and syntactic subjects: A study of Greek and Italian near-native speakers of English. International Journal of Bilingualism. 2004; 8:257–277. doi:10.1177/13670069040080030601.

Ulbrich C, Ordin M. Can L2-English influence L1-German? The case of post-vocalic /r/. Journal of Phonetics. 2014; 45:26–42. doi:10.1016/j.wocn.2014.02.008.

van den Bosch A, Daelemans W. Implicit schemata and categories in memory-based language processing. Language and Speech. 2013; 56:308–326. doi:10.1177/0023830913484902.

Walker A, Hay J. Congruence between ‘word age' and ‘voice age' facilitates lexical access. Laboratory Phonology. 2011; 2:219–237. doi:10.1515/labphon.2011.007.

Wanrooij K, Escudero P, Raijmakers MEJ. What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning. Journal of Phonetics. 2013; 41:307–319. doi:10.1016/j.wocn.2013.03.005.

Weatherholtz, K. Perceptual learning of systemic cross-category vowel variation. Ohio State University; Columbus, Ohio: 2015. Unpublished doctoral dissertation

Weatherholtz K, Jaeger TF. Speech perception and generalization across speakers and accents. submitted.

Weiner EJ, Labov W. Constraints on the agentless passive. Journal of Linguistics. 1983; 19:29–58. doi: 10.1017/S0022226700007441.

Weiss DJ, Gerfen C, Mitchel AD. Speech segmentation in a simulated bilingual environment: A challenge for statistical learning? Language Learning and Development. 2009; 5:30–49. doi: 10.1080/15475440802340101. [PubMed: 24729760]

Wells JB, Christiansen MH, Race DS, Acheson DJ, MacDonald MC. Experience and sentence processing: Statistical learning and relative clause comprehension. Cognitive Psychology. 2009; 58:250–271. doi:10.1016/j.cogpsych.2008.08.002. [PubMed: 18922516]

White, L. Grammatical theory: Interfaces and L2 knowledge.. In: Ritchie, WC., Bhatia, TK., editors. The new handbook of second language acquisition. Emerald; Bingley, UK: 2009. p. 49-65.

White L. Research timeline: Universal Grammar, crosslinguistic variation and second language acquisition. Language Teaching. 2012; 45:309–328. doi:10.1017/S0261444812000146.

White, L. Linguistic theory, universal grammar, and second language acquisition.. In: VanPatten, B., Williams, J., editors. Theories in second language acquisition: An introduction. 2nd ed.. Routledge; New York: 2015. p. 34-53.

Wonnacott E, Newport EL, Tanenhaus MK. Acquiring and processing verb argument structure: Distributional learning in a miniature language. Cognitive Psychology. 2008; 56:165–209. doi: 10.1016/j.cogpsych.2007.04.002. [PubMed: 17662707]

Wrembel M. L2-accented speech in L3 production. International Journal of Multilingualism. 2010; 7:75–90. doi:10.1080/14790710902972263.

Wunder, E-M. Crosslinguistic influence in multilingual language acquisition: Phonology in third or additional language acquisition.. In: De Angelis, G., Dewaele, J-M., editors. New trends in crosslinguistic influence and multilingualism research. Multilingual Matters; Bristol, UK: 2011. p. 105-128.

Yildirim I, Degen J, Tanenhaus MK, Jaeger TF. Talker-specific adaptation in quantifier interpretation. Journal of Memory and Language. 2015 Published online 11 November 2015. doi:10.1016/j.jml.2015.08.003.

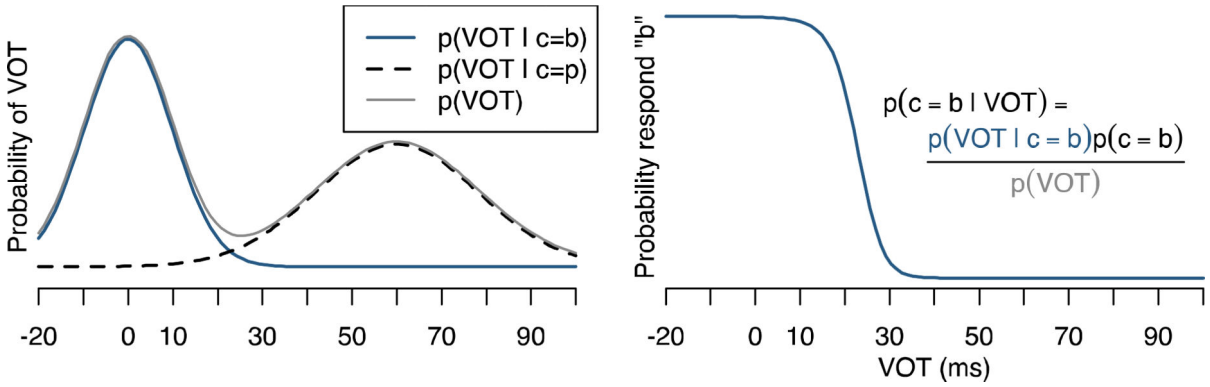

Figure 1.

Bayes' rule provides a link between the probability distribution over acoustic-phonetic cues given categories and the classification function. We illustrate this relation for the categories /b/ and /p/, and the voice onset time (VOT) cue, which is one of the primary cues to voicing in English. For a given VOT value, the probability that it corresponds to, say, a /b/ is proportional to the probability of producing that particular VOT value given the talker intended to produce /b/.

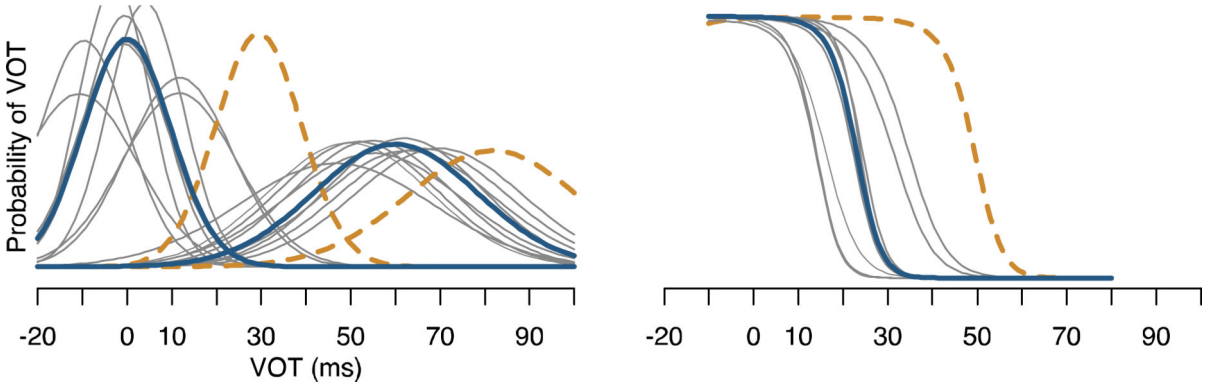

Figure 2.

Visualization of between-talker variability in /b/–/p/ production: distributions of voice onset time (VOT) values for /b/ and /p/ in English (left panel) and rational classification curves of sound tokens along the [b]–[p] continuum (right panel) given the distributions shown on the left. The depicted data are hypothetical but plausible (for comparison, see Allen et al., 2003).

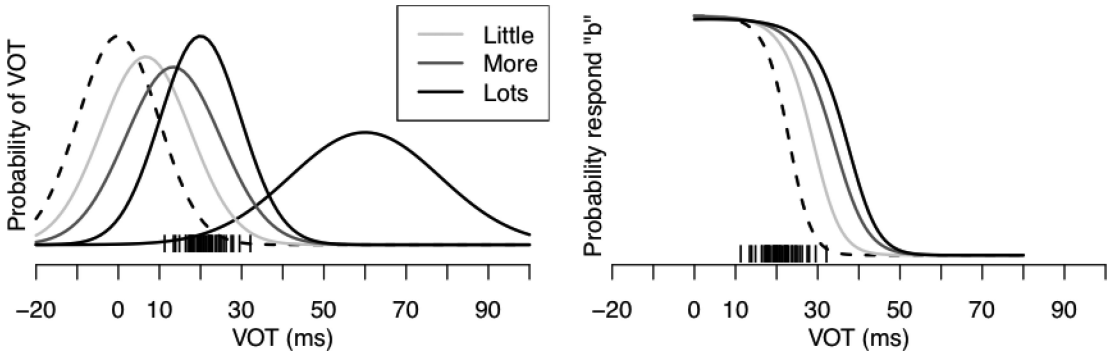

Figure 3.

Illustration of implicit statistical learning during perceptual recalibration (based on Kleinschmidt & Jaeger, 2015): changes to the beliefs about the category-specific cue distributions based on different amounts of exposure to the recalibration stimuli, shown as vertical dashes on the x-axis (left panel) and resulting changes to the classification function (right panel). A model based on the principles of Bayesian (or normative) inference provides a good fit against recalibration and other phonetic adaptation behavior (Clayards et al., 2008; Kleinschmidt & Jaeger, 2011, 2012).

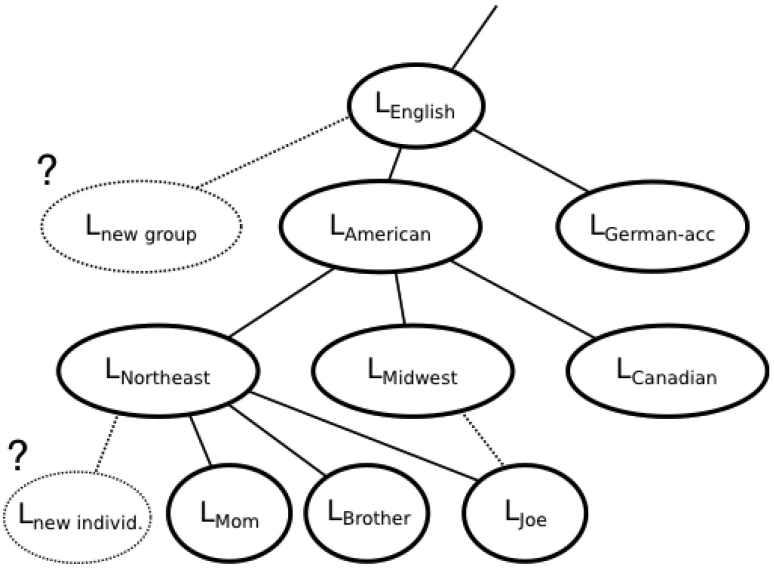

Figure 4.

Schematic visualization of a hypothetical listener's structured, uncertain beliefs about different language models (mini-grammars). Each node in the graph corresponds to a set of beliefs about language models. Dotted nodes/edges indicate uncertainty arising from the possibility of inducing new group or individual talker representations or reclassifying a representation (L_{Joe}) across levels.

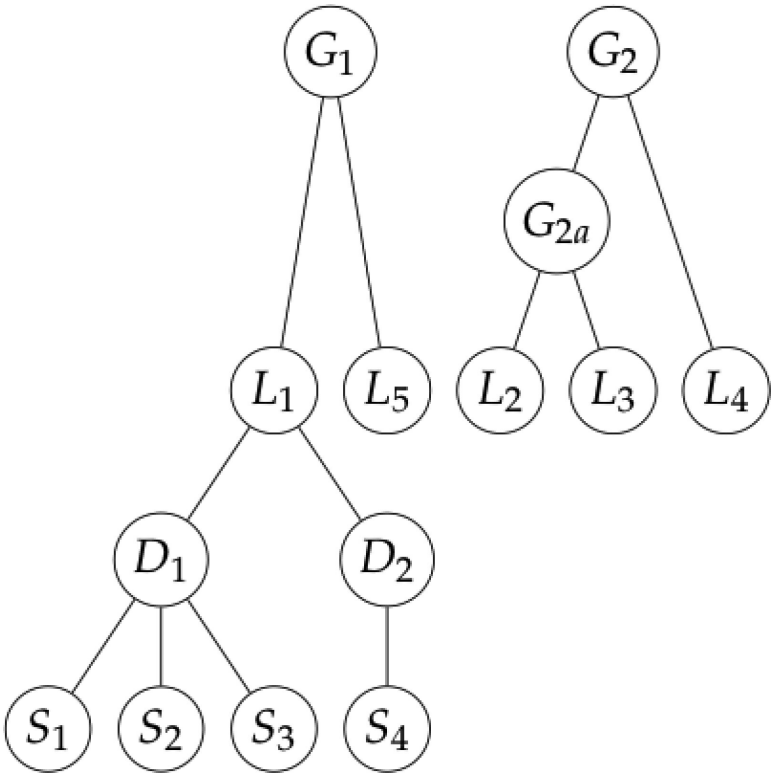

Figure 5.

An example of a multilingual environment, where languages, dialects, and talkers cluster based on similarity (L = language, G = language group, D = dialect, S = speaker). Language-internal structure is shown only for L1, but similar structures are present in all other languages. A specific example of this language environment is as follows: G1 = Germanic, G2 = Romance, G2a = Western Romance, L1 = English, L2 = Spanish, L3 = Italian, L4 = Romanian, L5 = German, D1 = American, D2 = Chinese-accented, S1 = Mom, S2 = Brother, S3 = Joe, S4 = Wei.

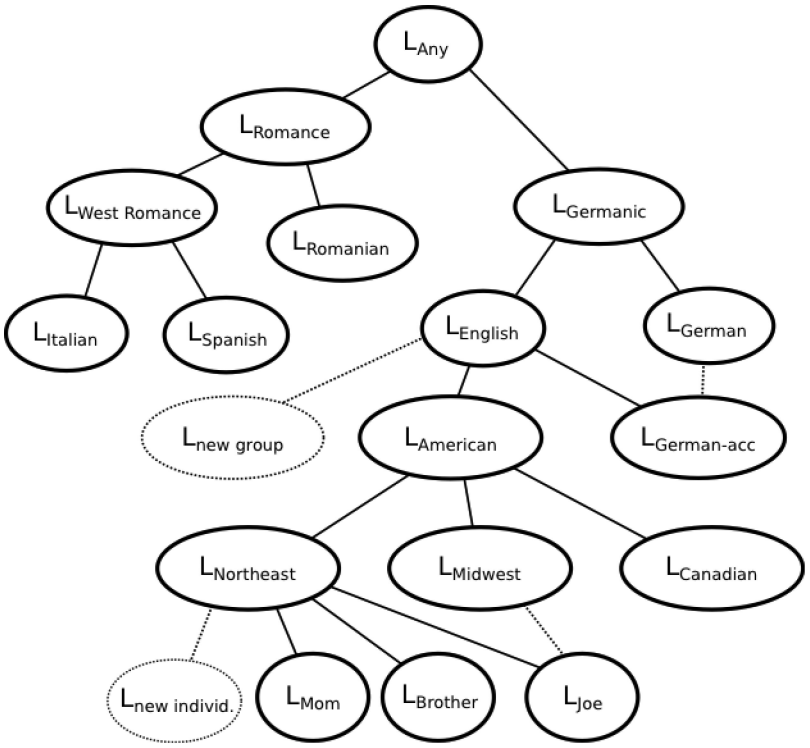

Figure 6.

Schematic visualization of a hypothetical listener's structured, uncertain beliefs about different language models, both within a single language (as shown for $L_{English}$) and across languages. Each node in the graph corresponds to a set of beliefs about language models. Dotted nodes/edges indicate uncertainty arising from the possibility of inducing new group or individual speaker representations or reclassifying a representation (L_{Joe}) across levels.