Frequency, Phonology, the Brain, and Second Language Acquisition: Laying the Ground Work for a Hybrid Approach

Scott Moisik

Department of Linguistics, University of Victoria
srmoisik@uvic.ca

ABSTRACT

A hybrid theory of second language acquisition is presented, which integrates postulates of Flege's (1992, 1995, 2003; Flege et al. 2003) Speech Learning Model and Bybee's (2001) proposals for Usage Based Grammar. As this proposal deviates from the Generative program, which does not permit frequency as an explanatory factor, a review is given of major currents in the literature concerning frequency in phonology and theories which incorporate frequency and patterns of distribution in phonological systems. The conclusion drawn from this review is that frequency is one of many performance factors worth considering as an explanatory factor in theories of phonology. The hybrid model is discussed and extended to a problem in second language acquisition of German by native English speakers. The SLM predicts that learners will not be able to establish a sound category for German dark-schwa, a sound which is remarkably *similar* to English schwa. A previous study shows that this is not only possible, but a production difference between the two German schwas was documented for relatively inexperienced speakers of the language. The hybrid model accounts for this exception to the SLM by appealing to the fact that the contrast between schwa and dark schwa in German is highly frequent.

Keywords: frequency; psycholinguistics; connectionism; emergentism; L2; German; dark-schwa; phonology; phonetics;

1 Introduction

The current 'received view' of linguistic research is the Generative model, which denies *performance* factors an explanatory role. While hybrid fields like psycholinguistics and neurolinguistics have already pushed deep into the frontiers of the brain and language processing, the traditional generative approach is still the predominant framework in which new linguists are trained and the context of mainstream linguistic research. Frequency represents one of many performance factors that have been systematically ruled-out of the generative paradigm in the interest of pursuing grammatical purity in the form of linguistic competence. The papers reviewed in this article strongly suggest otherwise; that by ignoring performance factors we risk distorting our view of language, placing it in a context isolated from everyday use and fluency. Frequency is central to this theoretical debate (e.g. Bybee & Hopper 2001; Ellis 2002) because of its salience as a concept for us. It relates directly to statistics and probability, both conceptual tools that are intuitive. It also is manifestly measurable; all one has to do is count. But the pith of this discussion is not to assert that frequency is the panacea to all of our linguistic woes. Rather, it is the starting point, or thread into a larger discussion concerning the future of mainstream linguistic research in a format that is "biologically, developmentally and ecologically plausible" (Ellis 1998: 640).

Although extensive research exists on the topic of how frequency impacts linguistic phenomena such as phonological representation (Frish 1996; Zuraw 2000), parsing (MacDonald 1994; Nimmo & Roodenrys 2002), phonotactics (Bybee 2001), first language acquisition (Curtin 2002; Nicoladis & Yin 2002), and the definition of language universals (Greenberg 1966), the generative framework has not acknowledged its explanatory role in linguistic research for the past forty years (Vihman 1996: 15-49; Ellis 2002: 175-77). Behaviorists conducted some of the early work on demonstrating a link between frequency and language. These theories lost favor in the domain of linguistics, however, when Chomsky published his 1957 critique of Skinner's *Verbal Behavior* (1957). This precipitated linguistic research with a focus on linguistic competence (Chomsky 1965) rather than performance or language use, a trend that is still prominent in the field. Today, with the increasing integration of linguistically oriented studies in cognitive (Langacker 1988), neurological, and computational programs, there has been a renewed interest in frequency and its effect on language (Ellis 1998). Of particular importance is the emergentist framework (Elman et al. 1996; MacWhinney 1998; Elman 1999); it has provided a theoretical foundation that integrates the concept of frequency among other neuro-cognitive and environmental factors which give rise to linguistic behavior. Emergentism, among other related contemporary frameworks such as connectionism, stands in stark contrast to the nativist program that has developed in tandem with Chomsky's Universal Grammar and Minimalist program (Cook 1988; Chomsky 1995). The perennial debate of *nature versus nurture* is once again raised by these opposed theoretical views. It should be noted, however, that the viewpoints contrast in how language is innate, rather than whether language is an exclusively internal or external phenomenon to the individual that engages in linguistic behavior (as far as the behaviorists were concerned: Watson 1924). The generative program is clearly aligned against the utility of linguistic performance facts as a means to obtain information about how language is represented. Their contention is that only by assessing linguistic competence do we learn about language in mind (Chomsky 1965). As we shall see, performance holds much more worth in models that afford frequency a role in explanation of linguistic phenomena. The ultimate goal of this paper is to present a selection of contemporary research being done that is reshaping the way that linguists think about language and contributing to a gradual paradigm shift (Ellis 1998, 2002) towards linguistic research that exhibits cognitive, neurological, ecological, and contextual validity.

2 The role of frequency

In two recent publications by Ellis (1998, 2001), concerns about the theoretical adequacy of the generative tradition are raised. The primary criticism is that generative linguistics has become hermetically preoccupied with the search for grammatical competence and in the process has ignored matters of "semantics, the functions of language, social, biological, experiential, cognitive aspects, [...] lexis, fluency, idiomaticity, pragmatics and discourse" (Ellis 1998: 634). Ellis is not a maverick in expressing these concerns either; Langacker (1988: 128) claims that excluding performance data severs the linguist from the very medium she is studying, severely jeopardizing the psychological validity of the research. Allen & Seidenberg (1999: 115-9) argue that, as a result of the competence-performance dichotomy, generative grammars have fatally limited their ability to explain linguistic phenomena. The main reasoning of their argument is that competence grammars are too abstract to be clearly related to performance factors (viz. Chomsky 1995: 380) and too exclusive of performance factors (such as memory capacity, perceptual and motor systems, and statistical information) to be cognitively, or neurologically valid. Ironically, however, it can be argued that the generalizations of competence grammars are never-the-less determined by metalinguistic judgments of speakers, which are known to be impacted by performance factors. Hopper (1987) asserts that language is not a static system uninfluenced by its day-to-day usage; Harris (1990) echoes the concern of treating language as a set of *a priori* conditions and ignoring the context and function associated with language. Elman (1999; Elman et al. 1996) along with neurobiologists (Ellis 1998) criticize generative grammar and UG for lacking plausibility in a neurological context, both in terms of inheritance and function. Pierrehumbert (2002) argues that statistical information (a matter of token frequency) about patterns can be viably associated with abstract phonological variables (such as morae and syllables). She asserts that statistical information allows us to "reach important conclusions about the nature of human language, conclusions that would elude us in a nonprobailisitic framework" (Pierrehumber 2002: 2). Such conclusions evidently tell us about the non-uniformity of phonological processes both diachronically and synchronically.

Ellis affords a large explanatory role to frequency; of particular interest here is the role frequency plays in phonology and phonotactics. Evidence from Frish et al. (2001) is raised to show how phonotactic judgements correlate with the frequency of a given phonotactic pattern; similar results are observed in Bybee (2001): judgments of word-likeness assigned by native speakers to nonce syllables correlate with the frequency of the phonotactic pattern in English, even when no phonotactic constraints are violated. Even more surprising is that nonce words with illegal syllable structure are regarded as more well-formed than nonce words with licit syllables that are infrequent, especially if the former contains a frequent suffix like *-ation* (c.f. /mɹu'pejʃn/ with /'splɛtɪsɑk/: see Coleman & Pierrehumbert 1997). Similar findings are demonstrated in infants (Jusczyk et al. 1994). The general claim is that humans possess the ability to perform distributional analyses on input stimuli; Ellis notes that this ability "is to be found in the plasticity of synaptic connections rather than abacuses or registers, but it constitutes counting nevertheless" (2001: 146). Thus, our sensitivity to the distributional patterns is not conscious, nor linguistic, but rather a property of our neurological organization. Even more tell-tale is the fact that numerous studies reported on by Ellis demonstrate an infant’s ability to rapidly process distributional information. In a syllable segmenting study by Saffran et al. (1996) infants exposed to an unbroken stream of syllables were sensitive to trisyllable groups that appeared as a unit given their increased frequency. Segmentation studies also document the importance of frequency in discovering the edges of linguistic units in an unbroken stream (e.g. Saffran et al. 1999). Beyond this, Ellis observes that humans are sensitive to distributional patterns on multiple planes, as we can integrate statistical information for multiple cues for word segmentation (Christiansen et al. 1998). In speech comprehension, comprehension is impeded when a low-frequency word must compete with high-frequency neighbors for activation (Lively et al. 1994).

Bybee (2001) presents several cases of diachronic phonological change that is explainable in terms of frequency. One of these processes is variable word-medial schwa-deletion in English. For example, common words like *every* are regularly pronounced without a medial schwa or its associated syllable slot (i.e. [ɛv.ɹi]). Mid-frequency words, such as *memory,* occur with a syllabic /r/, and rare words like *mammary* are produced with both a schwa and an /r/. Blevins (2004) proposes a model of diachronic sound change that incorporates frequency as a vital component. The model assumes that language change results from the discontinuous transmission and non-homogeneity of linguistic structure from person-to-person and generation-to-generation. Specifically, each speaker lexically encodes the phonetic variability of their langauge; language users determine the phonological form of words by the relative frequency of various competing phonetic forms (Blevins 2004: 41).

Pierrehumbert's Exemplar Dynamics (2001) and Probabilistic Phonology (2002) models are both built upon the assumption that phonological form is in part determined by the relative frequency of ambient phonetic tokens. In exemplar dynamics, phonetic variation for a given form (a phoneme or a word) is encoded into a cluster of the variants or exemplars encountered over the language user's lifetime. Thus, the relative frequencies of phonetic variants play a large role in shaping the content of the exemplar clusters. The resulting model is used to explain fine-grained phonetic patterns, the incremental changes observed in a language user's phonological patterns, and a diachronic class of sound changes involving lenition (such as the *every-memory-mammary* change cited above). Probabilistic Phonology is an attempt to render phonology more tractable to the observation that phonetic categories vary continuously and organically. This goes against the traditional generative assumption that phonologies are "(radically) underspecified" (Roca 1994: 53-62); in statistical models of language, the phoneme and lexical items contain "redundant and detailed" (Bybee 2001: 40) information. The correspondent of the phoneme in Probabilistic Phonology is the phonetic category, which is the locus of frequent linguistic output; the region which has the highest probability of being actually implemented by a given language. Pierrehumbert employs this model to discuss limitations on the language learner's ability to make inferences about phonetic categories as well as phonological constraints. The receptive vocabulary size and the volume of speech encountered on a day-to-day basis impact the distribution of the cues to these linguistic structures. The model provides an explanation for why vowel inventories do not exploit the entire acoustic space; probabilistically determined overlap among vowel tokens reduces the discriminbility of the system. Without gaps in the overall system - areas of low probability of vowel occurrence - the language would suffer from low discriminability.

Recently, facts of frequency have begun to appear in generative models and, while it is not explored fully here, it is worth noting. The traditional application of constraints in Optimality Theory (OT: Prince & Smolensky 1993) is categorical; constraint ranking is said to be held in a strict dominance formation. In this model, it is possible for constraints to be unranked with respects to one another, but in this does not conflict with strict dominance, rather it merely conflates the constraints. However, as proposed and employed in (Frish 1996; Hayes and MacEachern 1998; Boersma 1998; Boersma & Hayes 1999; Zuraw 2000; Escudero 2006), stochastic constraint ranking has been explored as a possible avenue of accounting for variability in output forms. In this model of constraint interaction, probability distributions define the likelihood of constraints being assigned a particular ranking in the grammar. At each attempted production, the grammar randomly selects a point in the ranking hierarchy to insert the constraint; crucially, this insertion point is constrained by the ranking's probability distribution. Hayes and MacEachern (1998: 48) claim that stochastic ranking predicts both output candidates in a similar fashion to classic OT, and allows predictions about output frequencies to be determined[1]. Thus, it would appear that the importance of frequency and related phenomena such as probability distributions are beginning to be re-acknowledged even by the presiding paradigm of linguistic research - the generative approach.

3 Theories that acknowledge frequency

The theories of language that incorporate frequency as an explanatory factor reviewed in this section notably reject the nativist hypothesis that linguistic structure is innate or hard-wired. This is one of the firmly entrenched assumptions that generative grammars make: despite the putative *poverty of the stimulus*, languages are learnable, because the principles and parameters of language are hard-wired into the brain. Thus, acquiring the grammatical structure of a language is a matter of finding evidence in the input for the parameter settings it employs. Or as Chomsky himself put it:

> A consideration of the character of the grammar that is acquired, the degenerate quality and narrowly limited extent of the available data, the striking uniformity of the resulting grammars, and their independence of intelligence, motivation, and emotional state, over wide ranges of variation, leave little hope that much of the structure of the language can be learned by an organism initially uniformed as to its general structure. (Chomsky 1965: 58)

The primary directive of generative grammars is implicit in the above passage: to understand the universal apparatus by which language acquisition is possible. Since the apparatus, referred to as Universal Grammar (UG), is robust enough to compensate for the poor sampling of language a prelinguistic child receives, frequency, among other factors is irrelevant to the linguists agenda. We shall see that the new approaches view language innateness in a radically different fashion.

Langacker's conception of grammar is formulated within the more general framework of Cognitive Linguistics (Rudzhka-Ostyn 1988), which pays heed to the psychological and cognitive environment that language occupies. The proposal in Langacker (1988) is for a usage-based model that incorporates frequency *ipso facto*: language is stored and represented as a "massive, highly redundant inventory of conventional units" (1988: 131). Unifying these units of linguistic storage are *schemas*, representations of linguistic patterns that emerge from the use of language through exposure and production. The statistical and distributional facts of language contribute to the influence that these schemas have on the grammar as a whole, and from speaker to speaker the grammar can exhibit variations that reflect the idiosyncratic linguistic experience of its user (1988: 130). The core tenets of usage-based grammar are diametrically opposed to those of generative grammar. In the later, grammars are assumed to be economically constructed sets of rules that generate the output of the grammar from a given input, and in this sense they are generative. Any information that is predictable is not listed in the grammar as it can be derived via linguistic rules. Langacker's usage-based grammar is described as "maximalist, non-reductive, and bottom-up" (1988: 131). Predictable details are stored in the grammar in addition to overarching patterns (which are referred to as schemas by Langacker). Thus, grammars are bottom-up in

[1] For further details on this matter the reader is encouraged to consult the original resource.

that the emphasis is on the details of linguistic forms rather than the rules that derive the forms. This reconception of grammar allows insight into the formation of grammatical principles/schemas. Instead of being a predetermined component of the linguistic system that becomes parameterized based on the ambient language, the language data gives rise to the pattern's extent in the grammar. Most relevant is the possibility for these patterns to encode frequency of the pattern by virtue of the number of tokens of a particular pattern that are encountered by the language user.

Bybee (2001) is taken to be representative of the basic outline of a usage-based grammar. She explores how the usage-based grammar is employed in the domain of phonology, and it will be useful to consider this aspect of her discussion. Six principals are described that characterize usage-based grammar; they are as follows. The first principle states that language experience impacts the way language is represented. One of the more important experiential effects is that of frequency: there is a tendency for higher frequency items to be represented more strongly. Strength is manifest in both accessibility of the form and the resistance to diachronic processes. The second principle is that language is not a modular system entirely distinct from other cognitive processes. As a consequence, there is no need in the system for predictable properties of language units (such as phonological representations) to be stripped away; the mind is capable of detailed storage. Thirdly, it is assumed that categorization is based on identity or similarity and that categorization of linguistic units operates in a manner analogous to other perceptual systems (such as vision). The fourth principle is that generalizations and patterns emerge from actual stored linguistic data in a manner similar to Langacker's (1988) schemas. Productivity of the grammar is described as a process of reference to existing forms; however, storage units can contain words that are fully inflected (i.e. /kæt + -s/ can be stored as /kæts/. The fifth principle advances Langacker's (1988) assumption that the grammar is redundant. In Bybee's model the grammar stores information at multiple levels of abstraction producing a highly redundant lexicon. The units of organization emerge from similarity across tokens and can be thought of in terms of traditional phonological units such as the syllable. The sixth and final principle is that phonology is assumed to be a form of procedural knowledge. This affords phonology a role in the production and decoding of linguistic constructions. An important implication of this idea is that phonology is thought to be learned in sets of procedural routines. These routines are categorizable and can be recruited in the production of similar words. As the degree of language use increases, the more the routines become automated and have the potential to reduce in form.

In usage-based models, principles of grammar are assumed to be created by and continually shaped through language use. Thus, there is no requirement for grammars to be prespecified for linguistic structure before the individual starts encountering language. Moreover, there is no disjunction between language and other cognitive functions, thus language can only be as 'prewired' as the entire cognitive apparatus. The question of innateness, however, still stands. Many questions are left unanswered by usage-based approaches with regards to how classification of linguistic units is carried out, or even begins for that matter. At times, a usage-based grammar will refer to structures being emergent. This is meant to describe a system where similarity in form gives rise to a generalization or abstraction over that form. Crucially, this is distinct from the nativist conception that the forms themselves are prewired and filter the perception of linguistic stimuli. The emergentist framework of language directly addresses the question of innateness, and thus will now be discussed.

Emergentism portrays language as a system in flux; it is fluid and adaptive; ultimately it is a system that has "structures [that] are unstable and manifested stochastically" (Bybee & Hopper 2001). Before a fruitful discussion of emergentism can be given, the issue of innateness must first be dealt with. Elman (1999) advances a taxonomy of innateness that contains three major divisions: representational innateness, architectural innateness, and chronotopic innateness. The goal of his analysis is to dispel the nebulous conception of innateness that has been perpetuated in previous literature (c.f. Pinker 1994). The first type, representational innateness, essentially embodies the traditional, generative perspective on the subject. Cognitive representations consist of patterns of synaptic connectivity; it is conceivable that the human genome could encode these patterns, and consequently linguistic principles and parameters. Studies on the genetic contribution to cortical development in humans and higher vertebrates indicate that plasticity, and not predetermination, is the nature of the mechanism (e.g. cortical plug transplants in vertebrates: O'Leary 1989). The consensus is that genes do not transparently contribute to complex behavioral patterns, even in organisms of a lower order of cortical complexity, such as the fruit fly (Greenspan 1995). Thus, Elman rejects representational innateness and along with it, the nativist hypothesis. The second type of innateness refers to the architectural constraints on neurons and neurological connections; for example, neuron firing threshold, neural interconnectivity, or macroscopic level connection amongst neural groupings. The third type of innateness is a matter of how developmental events are sequenced temporally. While this type of innate constraint on language is partly genetic in origin- genes set the gross schedule of neurological development -external factors play a role as well (Elman 1999: 5).

Thus, it is important to observe that emergentism does not dispense with innateness altogether. Rather, innateness is vastly reconceived, from its previous role as pre-specification of linguistic structure in order to bootstrap the learning process. Rather, grammar in emergentism is created through a conspiracy or confluence of forces acting on the individual. Two of these forces are the 'innate' or biologically determined constraints of neural architecture and developmental sequencing. The nature of the perceptual system and the form and distribution of tokens in the external stimuli constitute other factors that contribute to the emergence of language.

To demonstrate his claims, Elman (1993) presents a neural network simulation of L1 acquisition.[2] The overall task for the network was to correctly predict sequential dependencies in sentence strings. Part of this goal then required that the model assign syntactic categories to the input it received. The model was capable of being specified for how many words it could process at a given time, essentially simulating a 'working memory'. Affording the model a large working memory at the outset yielded mediocre results on the prediction task. However, imposing a developmental sequence on the model, where the size of working memory increased gradually over the training to handle more and more words at a time, yielded a vast improvement on the task. Elman concludes from this that by constraining

---

[2] This network begins with a series of network nodes (representing neurons) in layers; connections can formed across layers and the weighting (overall activation level) of these connections is influenced by the frequency of a particular input (see Nadel et al. 1989). Initially, all the weightings are randomly set so the network has no internal structure. Exposure to input allows the model to readjust weightings according to the types of patterns it is exposed to. These patterns can then be analyzed to identify what the model 'learned'.

the problem space, that is, by starting off with small chunks of grammar and gradually increasing the size of sentences that are analyzed, the learning task was made tractable. From this, Elman reasons that the very reason why a child is a successful language acquirer is not because she is endowed with preconfigured knowledge of grammar in the form of a Language Acquisition Device that later atrophies, but rather because their cognitive capacity initially allows them to process only small chunks of language.

4 Phonology in the brain

In an attempt to understand phonology from a neurologically and cognitively valid perspective, researchers must ask questions about the realization of phonological representations in the brain. The connectionist/ermegentist approach assumes that phonology is the result of an emergent set of representations mediating between perceptual and production components of the brain and the storage of semantic representations (e.g. Langacker 1988). Critically, the distribution of phonological data an individual encounters in their lifetime influences the structure of phonological representations. Thus, the content of representations includes redundant, predictable information and the frequency of variation in form is implicitly recorded by means of exemplar cluster density.

Bybee (2001) offers a conceptual reanalysis of the phoneme in light of current and previous research into perception and encoding of phonological units. The canonical approach to the psychological organization of phonemic units is that they have distinct boundaries and are categorically perceived. One criticism of this approach Bybee raises, is that often the underlying specification of phonemes is arbitrary, particularly where there are no morphological alterations to determine which phoneme ought to be used. An example of this is the problem with flaps in words like *butter* and *ladder*, or whether stops in /sC/ clusters are underlyingly voiced or voiceless. Another approach is to allow redundant and detailed information to characterize phonological representations. This assumption works hand-in-hand with the idea of exemplar dynamics (see Pierrehumber 2001) and provides the basis for a radical reconceptualization of the phoneme in a neurological context. As figure 1.0 shows, the phoneme is conceived of as a series of tokens distributed over a region in perceptual space. In this exemplar cluster representation, font size is used to represent the relative frequency of tokens in the phonemic clustering. In the case of English apical-alveolar consonants, the flap is represented as shared region in perceptual space between the /t/ and /d/ phonemes. Rather than discrete phonemic categories, Bybee argues for a phonological system built up of phonetic categories that store the set of exemplars representing that category. In this model, phonetic categories can be associated on multiple dimensions with other phonetic categories and non-phonetic categories such as context of use. The main problem is identifying how such a model can be realized neurologically.

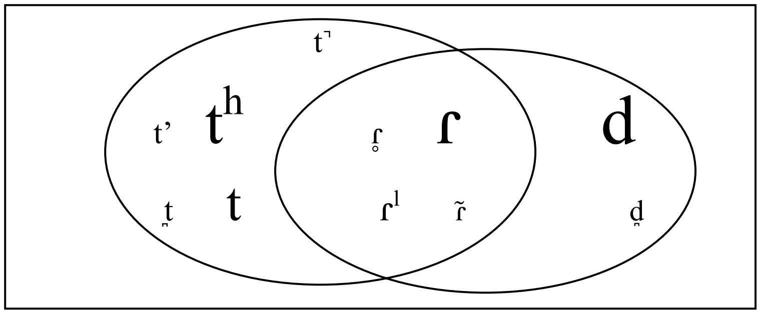

Figure 1.0 Schematic of overlapping exemplar cluster representation of phonetic categories for English apical-alveolar consonants; frequency and consequently representational strength is represented by font size: larger font indicates stronger representation/greater number of tokens (inspired byBybee 2001; Pierrehumbert 2001).

On the level of neural mapping of phonetic categories, Guenther & Gjaja (1996) demonstrated, through a neural network simulation, how auditory neural organization is dependent, in part, on the rate of occurrence of sounds in the individual's linguistic input. The assumption is that input stimuli is not stored specifically, but rather modifies the weighting of synaptic firing and does not hinge upon token labels or prewired knowledge of phonetic categories. Exposure to native language sound distributions generates non-uniformities in the firing patterns of neurological auditory units, providing a means to explain how infant perception gradually becomes categorical as more stimuli are encountered.

Neuroscience has provided us with a working knowledge of neurological organization, consisting primarily of neurons and the synapses they form with other neurons. While it is often claimed that psychology (and subordinately the subject of language in mind) should be free of neurological speculation because it is far too complex, connectionists argue that the task is made tractable through simulation of neurological architecture (Nadel et al. 1989). These models are structured on connective atoms, neuron-like processing units, intended to represent the neuron and its architecture as closely as possible. The use of these simulations offers insight into how a subsymbolic system (Smolensky 1997) might achieve symbolic representation through the use of connectives that can only interface mathematically with each other. Furthermore, these systems simulate rule-like behavior without being constructed on a rule framework (Ellis 1998: 638). While the topic of how neural network simulations function is important, focus will be placed on the findings of a number of these systems with respects to language, and more specifically phonology.

Allen & Seidenberg (1999) train a neural network to create connections between input forms (words) and their corresponding meaning (word-level semantic representations) to simulate comprehension. In addition, the converse mapping (i.e. meaning-to-form) is also implemented so that the network can produce utterances given an input meaning. By analyzing the pattern of network vector activation at the semantic layer of nodes, an assessment of its ability to comprehend words is demonstrated; performance on sentence comprehension is measured by comparing word level accuracy across an utterance. Similar tests were performed on its production abilities. Both tests reveal that the network was capable of dealing with novel utterances based on a limited training corpus. In addition, they provide evidence that the network possesses the ability to form grammaticality judgments, a metalinguistic ability, by means of a feedback loop between comprehension-level connections and production-level ones configured during training. Statistically significant behavior in network node activation is observed when novel ungrammatical sentences are presented to the network. As a whole, the network represents a dramatic demonstration of how patterns of grammaticality can emerge in a system made of simple connective units. Prewiring is not required to determine the pattern of connectivity; rather corpus distributions are sufficient to allow the network to identify patterns in the data.

Plaut & Kello (1999) discuss a neural network model of phonological development in the framework of emergentism. The authors seek to model how an infant surmounts the problems of phonetic variability and semantic arbitrariness to arrive at phonological representations of lexical items. The primitives of phonological and lexical structure are distributed throughout the system and, as acquisition takes place, emerge due to similarity in form of the input tokens. Thus, the phonological system, in their model, is presented as a tripartite link between acoustic input, semantic representations, and articulatory mechanisms. Phonological links to semantic representations emerge from repeated exposure to ephemeral acoustic input. Thus, the phonology serves as the memory required to map acoustic input onto semantic content in language comprehension; activation is persistent from one segment to the next as the acoustic input unfolds, until a semantic target is finally converged upon. The authors demonstrate how the model can both be said to have acquired a phonology that enables comprehension and production. This is done by analyzing the type of network connections that are formed across units representing various components of the model, such as that between acoustics and semantics. The extent to which the connections parallel the predetermined relationship between an acoustic form and its semantic feature bundle indicates whether the model converged upon the correct phonological representation. It is crucial to remember that the connections among nodes in the network are impacted by the distribution of forms in the input. While this model does not show how phonological patterns can be generalized to new forms, largely because there is no attempt to represent morphology, it is illustrative of how phonology must use frequency distributions to operate in a noisy system without any prewiring to aid it.

A final example of a neural network simulation applied to language comes from Elman (1990). The network is trained on a sequence of 'phonemes' (their orthographic counterpart) and tasked with predicting the next phoneme in a sequence when given an unfamiliar input string. By virtue of building and weighting network connections to represent the statistical dependencies observed in the training corpus, the network was able to perform the task with a relatively high degree of accuracy. Performance would initially be poor with the first segment, but gradually increase as more segments were encountered. When the incorrect activation pattern for a particular segment was made, often the network would still have activated the correct category of the segment (e.g. whether it was a consonant or a vowel). While this model is relatively simplistic (it is an early neural network simulation: c.f. Vihman 1996), it does offer insight into how simple phonological patterns, such as syllable structure, can emerge from mere exposure to a large corpus of data.

In summary, there is a rich body of research into the representability of language from a neurological perspective, assuming that language principles are not hard wired into the brain at birth. The models themselves are incredibly simple- they are only specified for a basic architecture of node layers- yet are still capable of performing complex linguistic tasks such as making grammaticality judgments. The lesson obtained from these results is that language does necessarily require innate representations in order to be a tractable object to represent neurologically. What is required is exposure to massive amounts of data and a system that is sensitive to the frequency distribution of forms in the input and capable of recording these patterns. The reader is encouraged to consult Ellis (1998) for further demonstrations of neural networks performing linguistic tasks.

5 Frequency & language acquisition/learning

In the emergentist/usage-based framework, first language (L1) acquisition is a matter of learning to use the language, rather than developing the knowledge required to make grammaticality judgments about one's own language (Allen & Seidenberg 1999: 120). Nor should language be regarded as instinctual; feral children may spontaneously learn to walk, but do not spontaneously generate a novel language in the absence of an ambient one (c.f. Lane 1979; Curtiss 1981). We are not born to use language, but can easily accommodate it, if sufficient input avails itself. Thus, from a neurological perspective, language learning can be accommodated by the brain in the same way that facial recognition is. Studies into categorical perception of infants (at four months old) demonstrate the brains ability to perform comparison of two stimuli that are highly similar (Eimas et al. 1971). In the event of exposure to language, infant auditory neurological systems encode aural stimuli and make connections to other systems (such as visuo-spacial processing and emotion: Ellis 2002: 655). A key feature of the input to this system is the distribution of forms; similar shades of phonetic features may occur more frequently than others. The following two studies provide insight into how auditory processing of linguistic information is influenced by distributional facts of the input.

While categorical perception is often brandished as the evidence that the phoneme is a psychologically real entity, current research into how distribution affects perception of phonetic contrasts forces us to re-evaluate how we conceive of the phoneme in a neuro-cognitive dimension. A study by Maye and Weiss (2003) reports on the discriminatory behavior of 8-month old infants, when exposed to different distributional patterns of tokens of a phonetic category (in this case Voice Onset Time: VOT). Two different types of distribution were employed: a bimodal one, and a monomodal one. Test group infants were exposed to tokens from the distribution for 2.5 minutes; the control group was presented with random tones. The experiment measured looking times as the dependent variable to determine whether a contrast was being perceived. Both groups (mono- and bimodal) were tested on the same phonetic stimuli, varying in VOT by 50ms. The results of their experiment show that exposure to a bimodal distribution, for even as little as 2.5 minutes, can prime categorical perception in infants. This indicates that the perceptual system exhibits a high degree of neural plasticity, and its corresponding representations are impacted by distributional information.

The hypothesis that L1 acquisition of phonemes is dependent on assessment of minimal pairs is challenged by Maye and Gerken (2000). The minimal pair hypothesis claims that an infant searches for and analyzes pairs of words minimally varying on a single phonetic category to infer what phonemes exist in the language (MacKain 1982). However, serious doubt is cast over this idea when one considers the receptive vocabularies of infants at the age that they start to display phonemic perception (Maye & Gerken 2000). Receptive vocabulary development begins at 12 months, but sensitivity to the ambient language's phonemic contrasts (the stage of perceptual reorganization) begins six months prior. In addition, incipient vocabularies rarely can be demonstrated to contain the extensive sets of minimal pairs required to allow such a phonemic analysis to be undertaken. Maye & Gerken attempt to validate a different hypothesis: phonemic acquisition is claimed to be the product of a perceptual reflex to sound distributions in the ambient language; this is known as the distribution-based hypothesis. Interesting support for this idea comes from experiments involving phonotactic pattern frequencies; children show a preference for high frequency patterns (e.g. *mubb*) over low frequency ones (e.g. *jurth*). Phonetic distributions of a language's phonemic categories are thought to be cognitively represented as nonuniformities in neuron firing patterns for a given phonetic category gradient (viz. the research of Guenther & Gjaja 1996). Such a model would thus account for, in a neurologically valid way, the behavioral flip-flop that occurs when infants begin to exhibit evidence of phonemic categorization and loss of non-native contrast perception.

Maye & Gerken (2000) test the distribution-based hypothesis by investigating the effect that exposure to a bimodal distribution on phoneme acquisition has, compared with exposure to a monomodal distribution. Adult participants were exposed to set of monosyllabic words from a pseudo language and then tested on their ability to detect whether words were different or the same in a pair-wise comparison. The results indicate that groups exposed to a bimodal distribution of the experimental contrast (/d/~/t/ pairs) treated it as a set of phonemes rather than a single category. The implication is that the bimodal group learned to identify minimal pairs without explicit training or semantic information. The findings of this study apparently run contrary to the idea that adults have diminished capacity for perceiving non-native contrasts. While it is true that English contains the very contrast being acquired in the experiment, this cannot be used as an explanation for performance of the bimodal group on the discrimination task because the monomodal group performed significantly differently on the same pairs in a monomodal distribution. These results set the stage for future research into L2 acquisition which incorporates frequency to be discussed in section 7.0.

6 Criticisms of frequency related accounts

To complete the review section of this article, I will give a presentation of criticisms against Ellis' article. The goal of this section is to anchor the discussion in a well-rounded discourse and to hopefully provide a common ground where the best of both approaches may be applied to a new hybrid model (see section 7.0).

The first topic under discussion is language innateness. Hulstijn (2002: 270) raises several concerns about the assumptions made by those who develop language learning neural networks (e.g. connectionists). The criticism is that these developers should not assume that the model is learning language from scratch as it has the architecture necessary to both perceive and store stimuli in a simulated fashion. Under the assumption that this accurately represents infant cognition, the claim is that we still cannot be exempt from talking about language as being non-innate, and thus the nativist hypothesis still prevails, albeit in an altered form. Indeed, Hulstijn raises a legitimate concern about the nature of cognition and perception; how is it that the developing human mind begins to perceive, store, and recognize sense-data from its environment. Neural network models do begin with this assumption, but, as has been discussed, this type of innateness is radically different than the type perpetuated in the UG framework. A detailed discussion of this has already been reviewed; i.e. the work of Elman (1999). Even Hulstijn concedes that infants "appear to possess the ability to count, through which they become subject to the powerful influence of frequency" (2002: 270). The problem of initial cognition is regarded as unresolved, but it is a matter to be decided by neurological and cognitive scientists as it concerns how sensory data is initially interpreted and stored by all of the perceptual systems of the body and not just the auditory system.

A further concern raised by Hulstijn is that implicit and explicit knowledge are impacted by frequency in different ways. Presumably, implicit knowledge, such as neural representations of auditory stimuli, is more subject to frequency effects by Hulstijn's account (2002: 271). The main point of his argument is that explicit knowledge about lexical form-meaning pairs responds to frequency in a manner that is different than implicit knowledge. Thus, exposure to a single token can lead to permanent acquisition of the word, and sometimes repeated exposure fails to trigger learning. These issues underscore the point that frequency is not intended to be taken as a panacea, used to resolve all issues of language learning. Rather, these examples indicate other psycho-cognitive phenomena such as *noticing* (e.g. Robinson 1995) and *neglect* (e.g. Mesulam 1981, 1990) may play a role, in addition to other possibilities.

The final comment Hulstijn makes concerning frequency comes from the case of L2 learners who have had more exposure to the L2 than their L1. These learners tend to demonstrate persistent and fossilized errors in morphosyntactic form pronunciation and use. Hulstijn's point is that despite the massive amount of L2 exposure these individuals receive, errors are not ameliorated as a frequency based account might lead us to suspect. Evidently, the case of L2 language is unique in that the L1 appears to inhibit the formation of additional grammatical and phonological knowledge. Hulstijn's concern is valid and serves as an additional factor that should be regarded in any theory attempting to explain L2 acquisition. At present, based on the discussion held in preceding sections, a possible conspiratory answer might be given to the question of how the L1 impacts the L2. Assuming that the neurocognitive system is most impressionable at infancy before any external input has been received, the impact of L1 data, and the frequency effects that correspond with it are considerable. By the time that infant becomes a child and matures in her language use, the neurological representations are firmly entrenched. These representations are hardened in the formative years having undergone neurological definition and pruning (Kosslyn 2005: 75) and are in play every time new stimulus is received. Thus, by the time a second language is encountered, the perceptual system must accommodate the new data through the old system, which will filter and distort the perception of foreign phonetic categories in a manner tantamount to the ideas expressed in the Native Language Magnet theory of L2 phonetic category acquisition (Guenther & Gjaja 1996; Iverson et al. 2003).

Gass and Mackey (2002) respond to Ellis (2002) by arguing that frequency is a complex matter and cannot be unequivocally applied to explanations of second language acquisition (SLA). Their intention, however, is not to rule out frequency as an explanatory factor in SLA, but rather to argue that it is a component of a confluence of forces acting upon learners. The first observation they make is that constructions that are rich in semantic content and frequent in the input, are nevertheless subject to relatively rigid placement in a developmental sequence. The example raised is that L2 learners acquire higher order question formation (e.g. Wh-questions with auxiliary inversion, negative questions, and tag questions: see Mackey 1999) late in the developmental sequence, despite their high frequency in the input. The authors also note, however, that frequency cannot be totally ruled out as there is evidence that grammatical constructions beyond the learner's current level are stored until the learner is ready to deal with them (Gass & Mackey 2002: 254).

Another criticism of frequency based accounts that Gass and Mackey (2002) raise is that ungrammaticality cannot be learned through positive evidence in the input; yet, it is clear that this type of knowledge is part of knowing a language. As only elements in the language that are actually instantiated can be countable, the claim is that frequency cannot be shown to play a role. On the other hand, they observe that the sheer absence of a particular construction in itself is a type of frequency information. Further argumentation supporting the generative view comes from observations made by White (1989) concerning an example of a linguistic complexity that can only be learned by being privy to innate knowledge of grammar. The particular construction is *wanna* contraction, which is banned in locations involving syntactic constituent extraction in between the two elements of the contraction (i.e. 'I want him to win the race'; c.f. 'Who do you want *t* to win?' vs. *'Who do you wanna win the race?'). The claim made by Gass and Mackey based on evidence such as this, is that mere language use subject to frequency effects could never lead to the type of abstract structure that is required to determine grammaticality. While at present there is no clear answer to this question, the work of researchers such as Allen and Seidenberg (1999) and Elman (1999) present cases of neural networks acquiring abstract grammatical structure. In the former, a neural network demonstrates the ability to perform grammaticality judgments on sentences with auxiliary inversion or tag questions, which involve movement and ellipsis, respectively. Elman's model is capable of employing the correct agreement on verbs even when a relative clause intervenes between the subject and the verb. While these cases are not as complex as those involving traces, they serve to demonstrate that modeling the acquisition of abstract syntax is not fully out of our reach.

7 A frequency-integrated approach to learning phonetic categories in SLA

The acquisition of second language (L2) phonemes has been a long standing matter of controversy. While many contend that the first language (L1) acts as a filter, the theoretical perspectives differ. From the perspective of phonologists, L2 learning is a matter of phonemic analysis of L2 phones into the phonological categories of the L1 (Brown 1997; LaCharité & Prévost 1999; Escudero 2006). The mechanism behind the phonemic analysis is the set of minimally contrastive features that drives the language learner's perception of foreign sounds. A classic case of this comes from Japanese (Brown 2000: 12): Japanese lacks a contrast between coronal approximants, unlike English, which has /l/ and /r/. Correspondingly, the minimally contrastive specification for the Japanese coronal approximant does not include the feature [coronal]. Given the lack of phonemic contrast of laterality in the language, English /r/ and /l/ are perceptually filtered into the same category. Consequently, the contrast is difficult to acquire and it is the learner's task to identify the set of features that drives the foreign contrast (c.f. Grenon 2006; Mah et al. 2006). The perspective of phonetically oriented studies runs contrary to the idea that features drive the perception of foreign sounds. Perception is fundamentally a matter of the psychoacoustic 'spaces' that an L2 learner carries with them from their L1. Thus, the ability to notice phonetic detail of a linguistically relevant contrast depends on the degree of perceptual similarity (viz. Flege's equivalence classification) to the L1 psychoacoustic space. An example of this comes from a case of differential substitution of French /y/ by English and Brasilian Portuguese (BP) learners (Rochet 1995). The English speakers tended to produce [u] for the sound, while the BP speakers produced [i]. It is well known that different languages possess different acoustic vowel spaces (Rochet 1995: 386; Johnson 1997: 102-7). Rochet argues that the learners hear the L2 sounds in the terms of their L1 phonological system by means of equivalence classification (see below for more details); accordingly, /y/ acoustically maps up with English /u/[3], while it is more close to BP /i/.

While both perspectives have presented compelling portrayals of the language learning process, a unified account is still lacking. The proposal made here is that a unified picture of L2 phonological acquisition will need to expand the scope of its research paradigm.

The critical area of research requiring elaboration is in the nature of how perception of the speech stream and phonemes map onto one another. Flege's Speech Learning Model (e.g. 1992, 1995, 2003) presents the idea that perception of phonetic categories is plastic and can respond to changes in the ambient language be it L1 or L2. Fundamentally, however, acquiring a new phonetic category is a matter of perceptual discrimination based on the active phonetic categories of the L1, and is subject to decline as a person ages. The mechanism driving perceptual equivalence classification is, at best, vaguely defined as a matter of acoustic similarity (Flege 1992: 572). A number of studies by Flege (1991, 2003; Flege et al. 2003) present evidence that L2 sounds are variably identified by L1 speakers as equivalent to an L1 sound. The success of acquiring a new phonetic category is argued to be dependent on the nature of this classification. Sounds representing a *new* phonetic category are acquired early on in the learning process because that particular category lacks competition from the L1 acoustic space. The difficulty lies in sounds that are *similar*, but not the same acoustically. The difficulty in perception lies in the fact that the phonetic category becomes absorbed into the pre-existing L1 acoustic space. This effect is exemplified by the above discussion of English and Brazilian Portuguese learners of French (Rochet 1995). In this case, the SLM would predict that acquisition of French /y/ would be difficult because both learner groups possess an acoustically *similar* L1 phonetic category (English /u/ and Brazilian Portuguese /i/).

This model is lacking in certain respects. First of all, the precise nature of perceptual mapping is poorly understood. Most studies approach the problem from an acoustic perspective, typically citing basic acoustic properties like formant structure, intensity, duration, and the like as grounds for comparison of L1 and L2 sound sets (e.g. Rochet 1995; Moisik 2006). In the Perceptual Assimilation Model (PAM: Best 1995; Best et al. 2001), mapping is a matter of both perceptual and articulatory degree of similarity; the degree of similarity can be determined by analyzing 'goodness-of-fit' judgments given by native speakers when assessing how much a foreign sound corresponds to an L1 phonetic category. Another interesting proposal was that psychoacoustic sensitivities can be modified by the phonetic categories of the L1. This is the position represented by the Native Language Magnet (NLM: Iverson et al. 2003) model. For example, native speakers of English show an increased sensitivity to changes in F3; this sensitivity is not attested for native Japanese speakers, who display an increased F2 sensitivity. Such observations lead Iverson et al. (2003) to conclude that the Japanese psychoacoustic map impedes the acquisition of the /l/-/r/ contrast in English. Flege's model forms predictions based on goodness-of-fit judgments

[3] Notably, /u/ tends to be quite fronted in its production in English.

(Flege 2003). While all of these different approaches to perceptual mapping provide insight into the nature of the problem, there remains the problem of what cognitive and neurological factors are involved.

There are also cases where the SLM fails to predict the behavior of L2 learners. A surprising result was obtained by Moisik (2006) where English learners of German managed to acquire dark-schwa, a sound classed as acoustically similar to English schwa (see below for details). The SLM would hold that such a contrast would not be attested by learners until much later in their language learning career.

It is proposed that, as Ellis (2002) argues, frequency should be incorporated into the research paradigm of second language acquisition. Currently, little work has been done on exploring SLA with usage-based or emergentist frameworks. Bybee's (2001) usage-based grammar (see section 3.0) encapsulates many of the ideas put forth by these two frameworks, and it will help to review it here. In her model, the concept of the phoneme is radically reevaluated in usage-based grammar. Bybee's argues convincingly for a type of 'radically overspecified' grammar (2001: 37-40), where redundant 'allophonic' information is actually incorporated into the representations of lexical units; the consequence of such a position is that the set of rules 'generating' the grammar is nearly non-existant. Rather, in Bybee's view, the process of speaking largely involves the activation of stored neuromuscular units associated with semantic material, as one produces the speech stream (2001: 15). In this perspective, the phoneme can be conceived of as a localized cluster of psycho-acoustically related units. The set of units that make up the cluster are referred to as exemplars, none of them exactly alike, but functionally classed according to the categorical exigencies of the ambient language. Evidence from first language acquisition studies show that distributional factors play a role in determining whether infants will perceive a phonetic contrast or not (Maye & Weiss 2003). The manifest behavior of individuals displaying categorical perception all but obscures the possibility that the representation of a single 'phoneme' could comprise the thousands of tokens an individual has been exposed to over their lifetime. Assuming Bybee's concept of the phoneme, the question is then, how can this re-conception of the phoneme integrated into second language learning. Bybee's model does not explore this possibility. The goal, here, is to outline a model that would unify the developments proposed by the SLM and similar models, with the type of work that has been done on frequency by Bybee (e.g. 2000, 2001) and others (e.g. Pierrehumbert 2001).

Thus, the goal of future research on this question should be to robustly outline the relationship between frequency and the perception of foreign phones. By unifying Bybee's and Flege's models, with the insights of emergentism also taken into consideration, this goal might be accomplishable. It is possible to synthesize the tenets[4] of the theories to come up with a set of predictions that would characterize the hypothesis of this new hybrid theory. They are as follows:

Prediction 1:

Frequency = strength of representation; the more a learner is exposed to second language input, the stronger the representations will be. Sounds that occur with a high degree of frequency will therefore be acquired faster than those that are occur with low frequency.

[4] See section 3.0 for a discussion of the principles of Bybee's usage-based model.

Prediction 2:

The phonetic/phonological representations of multiple languages are stored in the same neurological loci (contra-modularity; phonetic categories exist in common phonological space (Flege 1995: 239)); the representation of phonemes (in either language) can change over time as more phonetic exemplars are added to the cluster. As the L1 is more firmly entrentched[5] it will be more resistant to change in the face of new exemplars.

Prediction 3:

Similarity and difference (i.e. referring to Flege's *similar* and *new* equivalence classification scheme; 1995: 239) are definable on a neurological level: the center of mass for phoneme exemplar clusters defines the field in which a non-native sound can be determined to be more like the native sound compared to another non-native sound. This prediction is testable by means of a neural network simulation, where acquisition time of a non-native category is tested with varying degrees of similarity.

Prediction 4:

'Age of learning' effects are predicted on account of neurological plasticity and the ever increasing collection and reinforcement of native phonetic categories.

The troubling findings of Moisik (2006), alluded to above, serve as an illustration of how this hybrid approach can be applied to SLA, specifically focusing on *prediction 1*. In Moisik (2006) the production of dark-schwa (a German sound symbolized as [ɐ] in the IPA; found in words such as *bitter* [bɪtɐ] 'bitter') by English learners of German was examined and compared to productions of German schwa (in words such as *bitte* [bɪtə] 'please'). The prediction of this study was that the English speakers would not be able to produce a contrast between the sounds given their acoustic similarity (see Moisik 2006: section 3.3[6]; also see Moulton 1962: 37), but would rather produce something similar to English schwa for both. This prediction was made using the SLM as a framework: thus, dark-schwa was predicted to be classed as *similar* to English schwa and therefore difficult to acquire. The results, however, indicated that the speakers did make a contrast between the sounds. While the productions were not exactly native like, acoustic analysis revealed that all English learners of German produced dark-schwa and schwa differently.

In terms of the original SLM this finding is problematic because the SLM predicts that sounds classed as perceptually *similar* should be difficult to acquire (Flege 1995: 239). On the assumption that dark-schwa is perceptually similar to english schwa, the SLM fails to make the correct prediction. Dark-schwa was acquired, as evidenced by the production of contrast between dark-schwa and schwa, even though the learners were all relatively inexperienced with German (Moisik 2006: 52-5).

---

[5] Over time, as more and more exemplars populate a cluster, the pattern becomes resistant to change and less variable (Pierrehumbert 2001: 11).

[6] n.b.: dark-schwa is closer to English schwa than any other vowel in English ([ɑ] & [ʌ] were the next most similar English sounds).

With the hybrid theory, however, the behavior of the English learners of German actually confirms the first prediction: frequency of the L2 input amounts to stronger representations. There is ample evidence that dark-schwa is among one of the most frequent sounds in German. Dark-schwa is the standard production of nouns ending in <-er> like *Zauber* [tsawbɐ] 'magic', plurals such as *Kinder* [kɪndɐ] 'children', and the comparative morpheme /-əʀ/ as in *besser* [bɛsɐ] 'better' (Hall 1992: 101). It is also found as an off-glide when it follows a vowel (Hall 1992: 156), or in the frequently used prefixes *er-, her-, ver-* (pronounced [ɛɐ̆-], [hɛɐ̆-], and [fɛɐ̆-]: Hall 1992: 101). The contrast between schwa and dark-schwa in German is described as occuring with "high frequency and high functional load (Hall 1992: 101; Barry: 1995). Thus, it is reasonable to conclude that L2 learners of German will have ample input for dark-schwa. An explanation in the terms of the hybrid theory would amount to the following:

1) dark-schwa tokens are perceptually equated with English schwa, thus the sounds are similar (SLM).
2) the high frequency of dark-schwa strengthens the representation of the sound allowing it to form a unique phonetic category (usage-based approach).
3) therefore, despite similarity of the sounds, the frequency of the foreign target is high enough to overcome the *similarity* effect and allow for rapid acquisition sound production.

This illustration of how the hybrid approach to the SLM works represents the first steps towards incorporating knowledge about frequency into traditional models of phonetic and phonological acquisition in the context of second language acquisition. Further studies will reveal if frequency is sufficient to provide an explanation for interlingual behavior, or if other performance factors need to be integrated as well, such as "semantic basicness, salience, communicative intent, and relevance" (Ellis 2002: 178). Additionally, only *prediction 1* of the hybrid theory could be supported with the Moisik (2006) data; future studies are required to test the other three predictions, all of which require the construction of neural network simulations. The direction that these predictions point future research towards may allow us to begin to answer some of the difficult questions posed by Flege's SLM model: for example, how does mapping of L2 sounds onto L1 sounds actually work?

8 Conclusion

The hybrid approach presented in this paper only represents one possible framework for making linguistic research, and specifically second language acquisition, more answerable to questions concerning performance factors and the neurological implementation of language. Future research in this framework needs to focus on the representation of the phoneme and how it impacts exposure to foreign sounds. Neurologically, this task is, admittedly, extremely difficult, but neural networks are allowing us to conceive of how it might be accomplished (Nadel et al. 1989: 21; Plaut & Kello 1999). The hybrid approach makes predictions that open up new avenues to explore Flege's Speech Learning Model: for example, simulation of equivalence classification in a neural network model is feasible. Several 'test subject' networks could be created, all of which begin with the same predetermined weightings of network connections. These connections would represent the distribution of exemplars in the 'L1', i.e. the networks phonemic representations. Each network could then be presented with the same set of 'L2' phones that differed in some fashion from the average pattern of activation for any given L1 phoneme. To test the impact of frequency on L2 phoneme acquisition, a target L2 phone that would be classed as *similar*[7] to an L1 phone could be identified and its frequency in the L2 corpus could be varied for each network. At the end of a 'learning task', the patterns of activation for the L2 target sound in each network could be measured to determine to what degree frequency of the L2 target form played a role in its acquisition. The degree of similarity of the network's representation of the L2 target pattern to the actual L2 target pattern would indicate the extent to which it had acquired the phonetic category. An experiment such as this would provide insight into the way that equivalence classification operates on a neurological level and allow us to test the remaining predictions of the hybrid approach (*predictions 2 to 4*).

The goal of this article was to deal with the theoretical issues that arise when frequency is considered to be a relevant factor in linguistic explanation. It turns out that many of the assumptions about language learning, language change, and representation need to be re-evaluated. This re-evaluation is on-going and has been for more than a decade. We are also at an exciting time where questions of how neurological systems create symbolic representations are becoming more accessible to us. Modern computing power is enabling complex simulations of neural networks, and these are providing insight into how the *mind* comes to exist in the *brain* (Nadel et al. 1989: 18-22). Linguistic theories like emergentism and usage-based grammar are making connections between linguistic facts and neurological ones. The proposal here strongly echoes Ellis' (2002) argument, that frequency has a place in linguistic explanation, and more specifically in explaining second language acquisition. Frequency is not the entire picture, but it is certainly an important factor.

References

Allen, J., & Seidenberg, M.S. (1999). The emergence of grammaticality in connectionist networks. In B. MacWhinney (Ed.) *Emergentist approaches to language: Proceedings of the 28th Carnegie Symposium on cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Barry, W. J. (1995). Schwa vs. schwa + /r/ in German. *Phonetica* 52. 228–235.

Best, C.T. (1995). A direct realist perspective on cross-language speech perception. *Speech perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-language Speech Research*. W. Strange ed. York: Timonium, MD. 167-200.

Best, C.T., McRoberts, G.W. and Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustic Society of America* 109. 775-794.

Blevins, J. (2004). *Evolutionary phonology*. Cambridge University Press.

[7] In this case, similarity would be defined according to the details laid out in *prediction 3* of the hybrid model.

Boersma, P. (1998). *Functional phonology: formalizing the interactions between articulatory and perceptual drives.* The Hague, Holland Academic Graphics.

Boersma, P. & Hayes, B. (1999). *Empirical tests of the Gradual Learning Algorithm.* Manuscript, University of Amsterdam and UCLA.

Brown, C. (1997). *The Acquisition of Segmental Structure: Consequences for Speech Perception and Second Language Acquisition.* Doctoral Dissertation. Montreal, PQ: McGill University.

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald, (Ed.) *Second Language Acquisition and Linguistic Theory*. Blackwell. 4–63.

Bybee, J. (2000). The phonology of the lexicon; evidence from lexical diffusion. In M. Barlow and S. Kemmerer (Eds.) *Usage-Based Models of Language*. Stanford: CSLI.

Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, J. & Hopper, P. (2001). *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton de Gruyter.

Chomsky, N. (1965). *Aspects of a theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1995). *The Minimalist Program*. MIT Press, Cambridge, MA

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes* 13. 221–268.

Coleman, J. & Pierrehumbert J. (1997). Stochastic Phonological Grammars and Acceptability. *3rd Meeting of the ACL Special Interest Group in Computational Phonology: Proceedings of the Workshop*, 12 July 1997. Association for Computational Linguistics, Somerset NJ. 49-56.

Cook, V.J. (1988). *Chomsky's Universal Grammar: An Introduction.* Cambridge, Mass.: Basil Blackwill.

Curtin, S. (2002). *Representational richness in phonological development*. Doctoral dissertation, University of Southern California.

Curtiss, S. (1981). Dissociations between language and cognition: Cases and implications. *Journal of Autism and Developmental Disorders*, 11. 15-30.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science* 171. 303-306.

Ellis, N. C.(1998). Emergentism, connectionism and language learning. *Language Learning* 48, 631–64.

Ellis, N. C. (2002). Frequency effects in language processing: a review with implications for theories of explicit and implicit language acquisition. *Studies in Second Language Acquisition* 24, 143–188.

Elman,J. L. (1990). Finding structure in time. *Cognitive Science* 14. 179–211.

Elman, J. (1993). Learning and development in neural networks: The importance of starting small. Cognition, 48. 71–99.

Elman, J. (1999). The emergence of language: A conspiracy theory. In B. MacWhinney, Ed. *The emergence of language.* 1–27. Mahwah, NJ: Erlbaum.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). Rethinking innateness: A connectionist perspective on development. Cambridge, MA: MIT Press.

Escudero, P. (2006). A linguistic model of the interrelation between perceptual and lexical L2 development. In the *Proceedings of Generative Approaches to Second Language Acquisition*. May 28-30, 2006.

Flege, J. (1992). Speech learning in a second language. In C. A. Ferguson, L. Menn, and C. Stoel-Gammon, (Eds.) *Phonological Development: Models, Research, Implications.* 565-604. Timonium, MD: York Press.

Flege, J. (1995). Second language speech learning. Theory, findings, and problems. In W. Strange ed. *Speech perception and linguistic experience: Issues in cross-language research*. 233-277. Timonium, MD.: York Press.

Flege, J. (2003a). Assessing constraints on second-language segmental production and perception. In N. Schiller & A. Meyer (Eds.). *Phonetics and phonology in language comprehension and production: Differences and similarities.* Vol. 6, 319–355. Berlin: Mouton de Gruyter.

Flege, James & Carlo Schirru, and Ian R. A. MacKay. (2003b). Interaction between the native and second language phonetic subsystems. *Speech Communication* 40. 467–491.

Gass, S. M. & A Mackey. (2002). Frequency effects and second language acquisition: A complex picture? *Studies in Second Language Acquisition* 24. 249-260.

Greenberg, J. H. (1966). *Language universals*. The Hague, The Netherlands: Mouton.

Greenspan, R. J. (1995). Understanding the genetic construction of behavior. *Science American* 272. 74–79.

Grenon, I. (2006, May). Adults still have direct access to Universal Grammar: Evidence from the perception of a non-native feature contrast. Paper presented at the conference of *Generative Approaches to Second Language Acquisition*. Banff, Alberta.

Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America* 100. 1111–1121.

Hall, C. (1992). *Modern German Pronunciation: An Introduction for Speaker's of English.* Glasgow: Bell & Bain Ltd.

Hayes, B. & MacEachern, M. (1998). Quatrain Form in English Folk Verse. *Language* 74. 473-507.

Harris, R. (1990). *The Foundations of Linguistic Theory: Selected Writings of Roy Harris.* Nigel Love (Ed.). London: Routledge.

Hopper, P. (1987). Emergent grammar. In the Proceedings of the *Berkley Linguistics Society* 13. 139-57.

Hulstijn, J. (2002). What does the impact of frequency tell us about the language acquisition device? *Studies in Second Language Acquisition* 24. 269-273. Cambridge University Press.

Iverson, P., Kuhl, P.K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87(1). B47–57.

Jusczyk, P. W., Paul L., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language* 33. 630-645.

Johnson, K. (1997). *Acoustic and auditory phonetics*. Cambridge, MA: Blackwell.

Kosslyn, S. M. (2005). *Fundamentals of Psychology: the brain, the person, the world*. 2nd Edition: Rosenberg, R. S. (Ed.). Pearson Education Inc.

LaCharité, D. & Prévost, P. (1999). The role of L1 and teaching in the acquisition of English sounds by francophones. In A. Greenhill, H. Littlefield & Tano, C. (Eds.), *Proceedings of the 23rd Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press. 373-385.

Lane, H. (1979). *The wild boy of Aveyron*. London: Granada Publishing.

Langacker, R. W. (1988). A Usage-Based Model. In B. Rudzka-Ostyn (Ed.), *Topics in Cognitive Linguistics*, 127-161. Amsterdam: John Benjamins.

Lively, S. E., Pisoni, D. B., & Goldinger, S. D. (1994). Spoken word recognition. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* 265–318. San Diego, CA: Academic Press.

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes* 9. 157–201.

MacKain, Kristine S. (1982). Assessing the role of experience on infants' speech discrimination. *Journal of Child Language* 9. 527-542.

Mackey, A. (1999). Input, interaction, and second language development: an empirical study of question formation in ESL. *Studies in Second Language Acquisition* 21. 557–587.

MacWhinney, B. (Ed.). (1998). *The emergence of language*. Hillsdale, NJ: Lawrence Erlbaum.

Mah, J., Goad, H., & Steinhauer, K. (2006). The trouble with [h]. In the Proceedings of Generative Approaches to Second Language Acquisition. May 28-30, 2006.

Maye, J. & Gerken, L. (2000). Learning phonemes without minimal pairs. In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development*. 522-533. Somerville, MA: Cascadilla Press

Maye, J. & Weiss, D. J. (2003). Statistical cues facilitate infants' discrimination of difficult phonetic contrasts. In B. Beachley, A. Brown, & F. Conlin (Eds.), *Proceedings of the 27th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press: 508–518.

Mesulam, M. M. (1981). A cortical network for directed attention and unilateral neglect. *Annals of Neurology* 10. 309-25

Mesulam M. M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Annals of Neurology* 28. 597–613.

Moisik, S. (2006). *The L2 Phonetic Acquisition of Nuclear-/r/: The Case of English and German.* Honours Thesis. University of Calgary.

Moulton, W. (1962). *The Sounds of English and German: A systematic analysis of contrasts between the sound systems.* Chicago and London: The University of Chicago Press.

Nadel, L., Cooper, L. A., Culicover, P., & Harnish, R. M. (1989). *Neural connections, mental computation*. MIT Press, Cambridge, MA.

Nicoladis, E., & Yin, H. (2002). The Role of Frequency in Acquisition of English and Chinese Compounds by Bilingual Children. In B. Skarabela (Ed.), *Proceedings of the 27th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press. 441-52.

Nimmo, L. M., & Roodenrys, S. (2002). Syllable frequency effects on phonological short-term memory tasks. *Applied Psycholinguistics* 23. 643-659.

O'Leary, D. D. M. (1989). Do cortical areas emerge from a protocortex? *Trends in Neuroscience* 12. 401–406.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In Bybee, J. & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure*. John Benjamins, Amsterdam. 137-157.

Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory phonology VII (phonology and phonetics)*. Berlin: Mouton de Gruyter. 101–140.

Pinker, S. (1994). *The language instinct*. Penguin Books, London.

Plaut, D. C., & Kello, C. T. (1999). The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach. In B. MacWhinney (Ed.), *The emergence of language*. Mahwah: Lawrence Erlbaum Associates. 381–415.

Prince, A. S., & Smolensky, P. (1993). Optimality theory: Constraint interaction in generative grammar. *Technical Report* 2. Center for Cognitive Science, Rutgers University.

Robinson, P. (1995). Attention, memory, and the 'noticing' hypothesis. *Language Learning* 45. 283-331.

Roca I. (1994). *Generative phonology*. Routledge: London & New York.

Rochet, B. (1995). Perception and production of second-language speech sounds by adults. In W. Strange (Ed.), *Speech perception and linguistic experience. Issues in cross-language research*. 379-410. Timonium, MD.: York Press.

Rudzhka-Ostyn, Brygida. (1988). In B. Rudzhka-Ostyn, (Ed.), *Topics in Cognitive Linguistics.* Philadelphia: John Benjamins.

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

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition* 70. 27–52.

Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.

Smolensky, P. (1997). Connectionist Modeling: Neural Computation/Mental Connections. In J. Haugeland (Ed.). *Mind Design II*. Cambridge, Massachusetts. Bradford Book, MIT Press.

Vihman, M. M. (1996). *Phonological development: The origins of language in the child*. Oxford, UK: Blackwell.

Watson, J. B. (1924). *Behaviorism*. Chicago: University of Chicago Press.

White, L. (1989). *Universal Grammar and second language acquisition*. Amsterdam: Benjamins.

Zuraw, K. (2000). *Patterned Exceptions in Phonology*. Doctoral dissertation, University of California, Los Angeles.