Chad DeChant
Computation and the Brain
Fall 2018

What can Natural Language Processing learn from the brain?

Computer vision systems have recently seen incredible advances, from classifying images with near human accuracy to generating fake but entirely believable images. Their success is largely due to the development of convolutional neural networks, which were inspired by the primate visual system. These neural network have, in turn, been used to further study the primate visual system (Khaligh-Razavi et al, 2014) and continue to be tweaked and modified to more faithfully mimic the functioning of that system (e.g. Nayebi, 2018). In contrast to computer vision, computational approaches to natural language processing have not participated in a similar virtuous circle with studies of how the brain learns and processes language. Why not?

There are at least three broad reasons why findings about language and the brain have not yet had a similar galvanizing effect on natural language processing (NLP). First, we simply have less information about language and the brain than we do about vision. Much of what we know about human vision comes from extensive experimentation on primates and other animals. Since language is unique to humans, studying animals is only of limited use — for comparison's sake, for example. The types and volume of data we have about vision is simply unavailable for language because the invasive experimental techniques which are so useful in studying other primates are too dangerous to be employed on people.

Second, language may in some sense be more complicated than vision. Comprehension and production of language in the brain is entangled with many other brain systems and functions. Computer vision systems improved a great deal just from using somewhat brain-inspired convolutional layers in neural networks but there may not be such a simple and straightforward technique in the brain's processing of language that can be so quickly adopted by NLP practitioners. Insights about the brain and language may be on a larger scale than convolutional operators for vision, more about the larger architecture of the whole system than one component part. NLP in a machine learning context is usually pursued on its own, applied to a specific, narrow task such as translation. If language in the brain is inherently a more multi-

faceted and even multi-modal enterprise, it may not be possible for the traditional simple NLP tasks to benefit much from insights about the brain. NLP researchers may need to think bigger and develop more capable systems operating over multiple modalities if they are to import lessons from the way language operates in the brain.

Third, there is a significant disconnect between the research communities studying language in the brain and those approaching natural language processing tasks from a machine learning perspective. This disconnect, typical of the divides that come from academic specialization, means that the findings and methods from one field are largely unknown to those working in another. This area is perhaps even more divided than most because those working on language and its functioning in humans include many disciplines, the distinctions between which may not be clear to outsiders such computer scientists:

1. neuroscientists or neurolinguists studying how neurons, regions, and connections of the brain comprehend and produce language, typically using imaging or electrophysiological recordings of brain activity;
2. psycholinguists studying the psychological factors involved in language, typically using observation or interactive psychology experiments to find regularities and structure in people's use and comprehension of language;
3. biolinguists studying the "relationship between the genotypes and phenotypes responsible for explaining human language" (Raimy 2012, quoted in Martins and Boeckx 2016);
4. theoretical linguists describing the nature and structure of language in general and (in theory at least) as a biological phenomenon;
5. computational linguists modeling how humans learn and use language;

   (See Sedivy 2020 Chapter 1 for more on who studies language).

In order to in a small way help bridge the divide between research communities, we will describe a few aspects of research on language in the brain which might serve as useful inspiration to those working on computational approaches to NLP : (1) the way the brain breaks

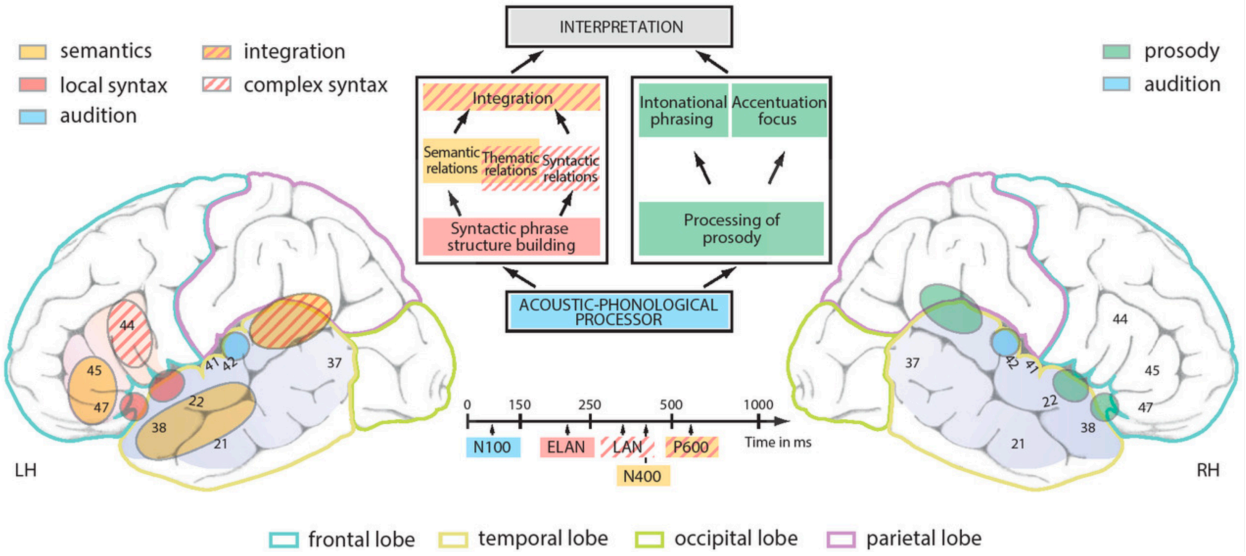
up language processing tasks in different brain areas and over time; (2) the order in which language capabilities are learned; and (3) theories of how lexical information is stored in the brain which suggest that it is linked to other sensory modalities and capabilities. At the end is an annotated bibliography of works cited and further readings.

I. Computation extended in space and time

A common goal in deep learning approaches to machine learning, including natural language processing, is an emphasis on training networks "end to end," beginning with as basic an input representation as possible and maximizing the likelihood of the desired output, e.g. a translation or parse tree. Backpropagation is left to structure the intermediate stages of processing without guidance from people whose "hand engineering" to define or break up a problem is viewed with suspicion. Although this approach has certainly been successful for the tasks it has been applied to, it is unclear whether it will scale to larger and more complicated challenges involving more difficult problems. It might be helpful instead to look at how evolution broke up the problem of processing language to see if that approach could be applied within natural language processing.

One model for understanding how the brain breaks up its work to understand spoken language is given in several papers and a book by Angela Friederici. A visual summary of her model is presented in the figure on the following page, borrowed from (Friederici, 2017). The areas involved in language comprehension in the left and right hemispheres are shown on the left and right, respectively. A rough outline of the time sequence of the illustrated processes is found in the center toward the bottom of the figure, while a flow chart illustrating the basics of her theory is found above it. Brain areas are labeled with their Brodmann area (BA) numbers.

The first step is to process the raw acoustic information coming from the ear, starting in auditory cortex (BA 42 in the figure above, shown in blue). This is thought to be undertaken in each hemisphere, with the left hemisphere focusing on recognizing speech sounds to classify them into phonemes and the right hemisphere processing longer duration so-called suprasegmental information involved in prosody (e.g. intonation, rhythm, and stress).



After sounds have been interpreted as parts of words in a language the listener understands, the next step is to match those sounds with particular words, preferentially in the left hemisphere. A word and information about it is abstractly thought of as being stored in the 'lexicon,' which includes information about its sound, meaning, and syntactic information (its part of speech and, if it is a verb, whether it takes a direct object and an indirect object).

Basic information about a word such as part of speech is accessed very quickly (110 - 170 ms after hearing a word), allowing for the beginnings of building up a phrase into a meaningful structure during a similar time window. Friederici is part of a school of researchers who believe the brain engages in "syntax-first" processing of sentences, in which syntactic and word category information alone plays a part in the initial combining of words into larger phrases. Other theories suggest that the meaning of words plays a more prominent role throughout the process. In a syntax-first approach, the meaning of words such as a determiner and a noun are not important for the simple process of combining or chunking them together into a local phrase structure constituent, in this case a so-called determiner phrase. In order to do this processing rapidly, the brain may use a template-matching procedure, matching incoming words to commonly used pattern templates, for example quickly beginning to construct a determiner phrase once a determiner is recognized. These early steps toward syntactic phrase structure building are localized to the anterior superior temporal gyrus and the frontal operculum, roughly in the areas marked in red in the figure.

Friederici sees one area serving a role both in the early building of local phrase structure and in the next phase, building up more complicated syntactic relations. This is BA 44, shown in the figure in red diagonal lines. It is here that she localizes a particularly important operation, combining two elements, a procedure called Merge. Merge is considered by some, including Noam Chomsky, to be the fundamental operation of syntax in human language. Once two elements are merged, that combined entity can then be merged with another constituent, *ad infinitum*, allowing for the construction and comprehension of arbitrarily complex sentences. Complex sentences are an additional burden on the brain, requiring the use of other areas of BA 44 as well as the posterior superior temporal gyrus/superior temporal sulcus. Areas carrying out working memory functions (left temporo-parietal cortex) may also be needed at this stage.

Nearly simultaneous with the build up of complex syntactic structures is further processing to make sense of semantic relations present in a sentence. As evident in the figure, this processing is widely distributed throughout the the left frontal, temporal, and parietal cortices. This is partly because, as we will see further in section III, semantic information is widely distributed throughout the brain, involving even areas not primarily associated with language such as visual or tactile sensory cortex and motor cortex. It is also because a lot of things are going on during this phase. An indication of the range of processing happening is given by studies of electrical signals in the brain, particularly an event related potential (ERP) known as N400. N400 is triggered when the brain detects a variety of violations of its semantic expectations, summarized by Friederici as:

> "(1) when a word does not have lexical status (i.e. a non-word or pseudo-word)
>
> (2) when the second word of a word pair does not fit the first word semantically
>
> (3) when in a sentence the selectional restriction of verb-argument relations is violated
>
> (4) when a word does not fit the preceding sentence context with respect to world knowledge or is simply unexpected." (Friederici, 2017)

Friederici localizes processes for basic semantic composition to the anterior temporal lobe, more complex composition to the angular gyrus, and strategic or "executive" semantic processes such as judging the similarity in meaning of two sentences to BA 47/45. Friederici advices those making psycholinguistic models to take account of these differences in function and location in

the brain — and those seeking to build functioning NLP systems to mimic those functions should consider doing the same.

Semantic and syntactic features are brought together nearly simultaneously for thematic role assignment, so called because the roles of nouns are assigned in the appropriate relations to a sentence's verb in order to spell out "who is doing what to whom." They are are again brought together in the next phase of processing, called integration, during which the brain gets a second chance to catch previously made mistakes. A well formed sentence might be reinterpreted or an ill-formed sentence may be mentally repaired to facilitate comprehension. The location of this stage in the brain is not yet fully pinned down, though one estimate is shown in the figure.

While much of the above work is predominantly carried out in the left hemisphere, the right hemisphere is also involved, and is particularly important for processing prosody. Such patterns of stress, pitch, and intonation are useful for syntactic processing and it is believed that communication between hemispheres allows this information to be shared with areas processing syntax.

II. Learning and brain development over time

One of the striking features of human language acquisition is how long it takes. The process begins even before birth, first words are only spoken after a child is at least one year old, major brain structures involved in language are only set in place by around age 10, and the learning of new words proceeds continuously until the teenage years and, though at a much slower clip, throughout a lifetime. While other systems, such as vision, continue to develop during childhood, it seems likely that the development of the language faculty is the most delayed in reaching full maturity.

According to Friederici, "although there are some variations in the speed of language acquisition across individuals, the sequence of the different development phases is invariant" (Friederici 2017). That the process of learning a language is so drawn out over time and that the sequence of stages in that process are the same across all language learners would suggest that there is something important about the process itself that those trying to develop

natural language understanding in computers would benefit from paying attention to. The stages are described in broad outline below.

Language learning begins in the womb, using the limited frequencies of sound that are able to be heard there. Newborns are born with language capabilities which are a mix of those they acquired in the womb, such as the melody of their mother tongue, and those which are genetically predetermined, such as being biased toward the perception of language-specific frequencies. Newborns only a few days old can distinguish between their mother tongue and another language, can pick out a different speech sound in a string of otherwise identical speech sounds, and have different brain responses to speech played forward and backward, suggesting they are already processing linguistic sound differently from non-linguistic sound. They also appear to be initially more responsive to longer-lasting suprasegmental (prosodic/melodic) information than segmental (phonological) information. (Friederici, 2017).

Babies next learn how to break up the incoming stream of speech sounds into individual words. Prosody and stress patterns within words are two tools babies use to find word boundaries. Perhaps surprisingly, they also use fine grained statistical patterns. In any given language, some syllable combinations are more likely to appear within words than between words. Babies can pick up on this to inform their guesses about where word boundaries are long before they have any idea what the words mean. An experiment done with 8 month old infants showed that they could use these patterns to begin to reliably identify individual words in an artificial language after listening to it for only two minutes (Sedivy, 2020, chapter 4).

The next step after being able to segment words is to learn their meaning, which infants can begin to do around the age of 6 months. A few months later they are able to generalize the meanings of the words they hear to form mental categories. Around the same time they begin to be able to chunk speech into phrases using prosodic clues like stresses and pauses. Syntactic categorization of words then follows as early as 18 months of age. When they are 3, children begin to analyze syntactic relations between nearby words and phrases. (Skeide and Friederici, 2016).

While their ability to produce and comprehend language is developing steadily throughout childhood, significant underlying changes are occurring in their brains to make this

happen. Interestingly, language processing in children is quite different than that in adults, even at ages (e.g. 7 or 8) when they are clearly capable of complex language comprehension and production. While we saw in the previous section that semantic and syntactic processing occur largely in separate areas of the brain in adults, this separation does not occur in children until they are 9 or 10 years old. Another significant difference is that some important connections between regions do not fully develop for many years. Of particular note is the dorsal pathway connecting the posterior superior temporal gyrus and BA 44, which is not mature until a child is 10 years of age. This pathway is considered critical for complex syntactic processing, especially of unusual sentence structures.

The gradual development of language processing ability in humans briefly outlined above has no parallel in machine learning approaches to NLP. The learning in machine learning is of an entirely different character. In humans, the architecture guiding and shaping the learning process changes over time, presumably in ways that support and are important for that learning. In machine learning, a fixed architecture is exposed to vast amounts of data and expected to learn with no significant changes to its basic structure. Even evolutionary approaches to designing deep learning networks are employed not to change their structure during learning but before learning begins.

III. Multi-modality and the development of lexical information

Many approaches to NLP from a machine learning approach involve learning representations — representations of words, sentences, features of a model, etc. There are two drawbacks to the way these are learned when viewed from a perspective informed by humans' language learning.

First, like the problem we just saw in section II, what we might call the space of the representations does not change over time. Children learn semantic concepts, how to represent them in the brain, and how to relate them to other concepts very slowly, beginning in the simplest possible terms — perhaps just positive and negative affect, for example. Then over time as their experience grows and they learn new ways that their prior concepts relate to each other, they must be able to effectively grow the representation space that these concepts are

manipulated in. For example, we saw that it is believed that syntactic information about words is somehow stored in a person's lexicon, yet we also saw that children show no evidence of appreciating even simple syntactic categories until they are almost 2 years old. So, unless the dimensions of lexical representation are genetically hard wired, at some point children's brains must be able to grow the representation space of the lexicon to make room for this newly appreciated syntactic dimension of words.

This kind of learning is not possible in today's machine learning techniques. As a language model, for example, learns to represent words as vectors, the learning does not alter the dimensionality of the vectors as the learning proceeds, e.g. starting from smaller vectors for simpler concepts when it has limited knowledge to embed in the vectors and slowly adding new dimensions to account for new sorts of information. Instead, all the vectors are the same size and remain so. Representations for all words are learned effectively at the same time, in stark contrast to humans' language learning. Whether trying to follow a more human-like course of learning would make for better NLP systems is an open question, but one worth trying to answer.

A second potential limitation of representations as learned by today's machine learning algorithms is that they are learned and used separately for different modalities. Word representations are learned from co-occurrence patterns in language corpora. Visual representations are learned from exposure to large numbers of images. Tasks like navigation through an environment or playing video games have, in turn, their own distinct representational spaces dependent on how their task has been structured. It is not like this in the brain. As we briefly saw in section I, there is evidence to believe that semantic representations are widely distributed across the brain and that representations primarily associated with one modality are not restricted to that modality but are instead shared and available to be combined with others from different modalities. For example, color information associated with the word "red' or "apple" would be stored in visual cortex, while motor information needed to make sense of the word "kick" would appear in or near regions of motor cortex engaged when kicking something. Exactly how semantic representations are shared across modalities is an intensely debated subject, touching on larger disagreements about how embodiment affects cognition more broadly

and referencing longstanding philosophical debates, all of which we will have to leave unaddressed.

There are many theories of semantic representation but, simplifying the four-fold taxonomy put forth by Meteyard et al (2012), we will contrast three types. The first sees lexical semantic representations as completely disconnected from other modalities' representations. Much like contemporary machine learning representations, words derive their meaning from their relations to other words. Meteyard at al refer to such theories, in which other modalities play no role in lexical semantics, as unembodied; the language system's finding itself operating in a body that is seeing, feeling, and acting simply does not influence its semantic representations. At the other end of the spectrum, theories of strong embodiment see other systems as entirely and inherently wrapped up in semantic processing. As Metayard et al write of these theories, "low level sensory and motor information is activated in primary cortical areas as part of routine semantic processing. This effectively pushes semantics out into primary cortical areas and makes it completely dependent on sensory and motor systems." Processing language results in "full simulation" of whatever a sentence describes; if a sentence involves walking over wet grass, motor cortex systems involved in walking will be engaged, perhaps using so-called mirror neurons, as will somatosensory cortex areas to simulate the feel of dew drops and grass blades breaking under foot.

Between the two extremes of unembodied and strongly embodied theories are a range of options, which we will collectively refer to here as intermediate embodiment. In a theory of intermediate embodiment, semantic representations in the lexicon are linked to sensory-motor content but not limited to them. On one kind of intermediate account, the semantic representations are themselves amodal and are based in an amodal "hub" which is connected to the underlying sensory and motor representations. Some researchers propose that such a hub is likely to be located in the anterior temporal lobe (ATL), based partly on the experience of people with brain lesions there and associated semantic disorders (Patterson and Lambon Ralph, 2016). The amodal representations are influenced by and in turn influence the representations associated with the individual modalities. On another intermediate account, "semantic representations are at least partly constituted by sensory-motor information" (Meteyard et al 2012). Some words may

have amodal components in their representations while others will be constituted by mixtures of individual modalities' representations.

Whether an intermediate or strongly embodied conception of semantic representation is more accurate, trying either as richer representations for NLP tasks, particularly those like image captioning which are inherently multi-modal, would seem a fruitful endeavor.

We have reviewed some areas of research into how the brain processes language and suggested preliminary connections to NLP work. The above brief forays into neuroscience and psycholinguistics are but glimpses of only a handful of models which themselves draw on only a subset of findings about language in the brain. This might sound daunting to even the most interested NLP researcher. Hopefully it points to the diversity of options for augmenting current NLP approaches and the room that exists for experimentation.

ANNOTATED BIBLIOGRAPHY (IN PROGRESS)

Khaligh-Razavi, Seyed-Mahdi, and Nikolaus Kriegeskorte. "Deep supervised, but not unsupervised, models may explain IT cortical representation." *PLoS computational biology* 10.11 (2014): e1003915.

Martins, Pedro Tiago, and Cedric Boeckx. "What we talk about when we talk about biolinguistics." *Linguistics Vanguard* 2.1 (2016).
Covers the history and current use of the term "biolinguistics." Although many view the term as synonymous with generative linguistics or specifically Chomsky's recent work on the Minimalist program, Martins and Boeckx argue for a broader understanding of biolinguistics as the study of the biology of language generally, e.g. genetics and neuronal dynamics.

Meteyard, Lotte, et al. "Coming of age: A review of embodiment and the neuroscience of semantics." *Cortex* 48.7 (2012): 788-804.
Reviews different theories of semantic representation, arranging them in four categories depending on where they fall on a spectrum of 'embodidness', i.e. how important the notion of embodied cognition is to each theory. The four categories of theory are: (1) unembodied theories which consider all semantic representation to be amodal with words depending only on their relations to other words for their meaning; (2) "secondary embodiment" theories which propose that semantic representations are basically amodal but are linked to modality specific representations; (3) "weak embodiment" theories in which semantic representations are "at least partly constituted by sensory-motor information," including abstract layers of representation that reside near the primary sensory areas; (4) "strong embodiment" theories in which low level sensory and motor information is needed for routine semantic processing.

Nayebi, A., Bear, D., Kubilius, J., Kar, K., Ganguli, S., Sussillo, D., DiCarlo, J.J. and Yamins, D.L., 2018. Task-Driven convolutional recurrent models of the visual system. In *Advances in Neural Information Processing Systems* (pp. 5290-5301).

Patterson, Karalyn, and Matthew A. Lambon Ralph. "The hub-and-spoke hypothesis of semantic memory." *Neurobiology of language*. Academic Press, 2016. 765-775.
The authors propose that semantic knowledge in the brain is organized in modality-specific regions ("spokes") all linked to a "transmodal" hub which contains "modality invariant" representations. The spokes contain such things as "the color of a camel in color regions, its shape in visual-form regions… its name in language-specific cortex." The hub "codes semantic similarity structure and represents concepts in a manner that abstracts away from the specific features of how they look or sound," etc.
Puts forth evidence to conclude that the hub resides in the anterior temporal lobe.

Skeide, Michael A., and Angela D. Friederici. "The ontogeny of the cortical language network." *Nature Reviews Neuroscience* 17.5 (2016): 323.
Skeide and Friederici trace the development of language comprehension and production in the brain. They propose a two phase model in which language learning is thought of as having a first, "bottom-up", phase primarily located in the temporal cortices and the ventral language network and a second, "top-down" phase involving the frontal cortex and dorsal network. Bottom-up processes are automatic and involve first processing and segmenting speech sounds into word forms, then accessing the meanings of those words. Some basic syntax is also present in the first few years but more complex syntactic processing requires the top-down phase.

BOOKS

Sedivy, Julie. *Language in mind: An introduction to psycholinguistics, Second Edition*. Oxford University Press, 2020 (sic).
Chapter 1
Box 1.2 gives a nice taxonomy of the different kinds of researchers studying human language

Chapter 2: Origins of Human Language
A basic introduction to some foundational questions in the study of language, such as defining characteristics of language (e.g. "*productivity:* the ability to use known symbols or linguistic units in new combinations"), what differentiates human language from the communication of other animals, attempts to teach nonhuman primates language, and sign language.
Summarizes some work on the evolution of human language, including arguments for and against a universal grammar, and the theories of Michael Tomasello on "the social underpinnings of language," arguing that joint attention — "the awareness between two or more individuals that they are paying attention to the same thing" — is a necessary prerequisite for the evolution of language in humans and its normal development in a child. [This work is of potential interest to those working on modeling the evolution of language in simple agents in a reinforcement learning setup.]

Chapter 3: Language and the Brain
Begins with a history of efforts to understand how and where the brain processes language, then moves into a quick and fairly high level introduction to the current state of the field. Techniques used to study the brain, including imaging and electroencephalography (EEG), are presented along with descriptions and illustrations of the functional neuroanatomy of language and a selection of key ERP components.

Chapter 4: Learning Sound Patterns

An overview of how the sounds of a language are learned and produced as well as the methods used by psychologists to study babies' understanding of language.
Of particular interest is the section on infants' learning to segment speech into discrete words, which they can learn to do for artificial languages after even very short exposure to the new language. It is thought they do so by learning to estimate the probabilities that a particular syllable will be followed by another; transition probabilities between syllables within a word are higher than those between neighboring words.

Chapter 5: Learning Words

Looks at how and when young children learn words from a psychology perspective. Children have some learning biases that help them pick up words, e.g. the whole object bias: "the assumption that a new word heard in the context of a salient object refers to the whole thing and not to its parts, color, surface," etc.
Children do not learn words simply by being exposed to them in the presence of the relevant objects or actions; the learning process depends on a great deal of social interaction and depends on their ability to infer the intent of their interlocutors.
How children learn to differentiate regular vs irregular verbs is also discussed, along with computational models of that learning.

Chapter 6: Learning the Structure of Sentences

Introduces some basic ideas about syntax, theories about how children learn syntax, and evidence for and against those theories. Rule-based theories of syntax maintain a "sharp boundary between memorized lexical representations and abstract rules that combine units in a compositional way" whereas constructionist accounts of syntax reject "the notion of a strict separation between memorized lexical items and combinatorial procedures, and relies instead on structural templates that combine abstract information with detailed information regarding specific words or phrases."

Chapter 7: Speech Perception

Lays out some of the difficulties involved in speech perception and some theories about what goes into it. The speech sounds we make and hear vary in a lot of ways depending on context and the environment we're speaking in. Sounds we hear vary continuously between different phonemes but we perceive phonemes in fairly rigid categories, hearing a sound as entirely one or another phoneme, not the mix it may actually be.
According to the motor theory of speech perception, there is a link between the perception and articulation of words and that "the perception of speech sounds involves accessing representations of the articulatory gestures that are required to make those speech sounds."

Friederici, Angela D. *Language in our brain: The origins of a uniquely human capacity*. MIT Press, 2017.

Chapter 5: The Brain's Critical Period for Language Acquisition

Language learned as a second language after a critical period is not learning the same as the first language. Up to age 3 learning a second language is "native-like for both syntax and semantics… After the age of 3 years acquisition is native-like for semantics but already non-native-like for syntax and … after puberty a second language does not seem to be acquired in a native-like manner." Some researchers hypothesize that the critical period is related to the delayed myelination of axons linking language-relevant regions of the brain (BA 44 and the posterior temporal cortex), suggesting that myelin inhibits the sprouting of neurons in the region where axons terminate, thus inhibiting learning.

Imagery and ERP patterns of brains processing sign language are very similar to processing spoken language, suggesting a "universal neural language system largely independent of the input modality."

Chapter 6: Ontogeny of the Neural Language Network

Traces the sequence of language learning stages in children which, despite some individual variation in timing, is the same for all. Builds on the model developed in Skeide and Friederici, 2016.