Psycholinguistics and Language

Lawrence Marsh

PSYCHOLINGUISTICS AND LANGUAGE

PSYCHOLINGUISTICS AND LANGUAGE

Lawrence Marsh

BIBLIOTEX
Digital Library

Psycholinguistics and Language
by Lawrence Marsh

www.bibliotex.com

Ebook ISBN: 9781984665300

Published by:
Bibliotex
Canada
Website: www.bibliotex.com

Contents

Chapter 1

Areas of Study

Language Acquisition

Language acquisition is the process by which humans acquire the capacity to perceive and comprehend language (in other words, gain the ability to be aware of language and to understand it), as well as to produce and use words and sentences to communicate.

Language acquisition involves structures, rules and representation. The capacity to use language successfully requires one to acquire a range of tools including phonology, morphology, syntax, semantics, and an extensive vocabulary. Language can be vocalized as in speech, or manual as in sign. Human language capacity is represented in the brain. Even though human language capacity is finite, one can say and understand an infinite number of sentences, which is based on a syntactic principle called recursion.

Evidence suggests that every individual has three recursive mechanisms that allow sentences to go indeterminately. These three mechanisms are:*relativization*, *complementation* and *coordination*.

There are two main guiding principles in first-language acquisition: speech perception always precedes speech production, and the gradually evolving system by which a child learns a language is built up one step at a time, beginning with the distinction between individual phonemes.

Linguists who are interested in child language acquisition have for many years questioned how language is acquired. Lidz et al. state "The question of how these structures are acquired, then, is more properly understood as the question of how a learner takes the surface forms in the input and converts them into abstract linguistic rules and representations."

Language acquisition usually refers to **first-language acquisition**, which studies infants' acquisition of their native language, whether that be spoken language or signed language, though it can also refer to **bilingual first language acquisition** (BFLA), which refers to an infant's simultaneous acquisition of two native languages.

This is distinguished from *second-language acquisition*, which deals with the acquisition (in both children and adults) of additional languages. In addition to speech, reading and writing a language with an entirely different script compounds the complexities of true foreign language literacy. Language acquisition is one of the quintessential human traits.

History

Some early observation-based ideas about language acquisition were proposed by Plato, who felt that word-meaning mapping in some form was innate. Additionally, Sanskrit grammarians debated for over twelve centuries whether humans' ability to recognize the meaning of words was god-given (possibly innate) or passed down by previous generations and learned from already established conventions: a child learning the word for *cow* by listening to trusted speakers talking about cows.

Philosophers in ancient societies were interested in how humans acquired the ability to understand and produce language well before empirical methods for testing those theories were developed, but for the most part they seemed to regard language acquisition as a subset of man's ability to acquire knowledge and learn concepts.

Empiricists, like Thomas Hobbes and John Locke, argued that knowledge (and, for Locke, language) emerge ultimately from abstracted sense impressions. These arguments lean towards the "nurture" side of the argument: that language is acquired through sensory experience, which led to Rudolf Carnap'sAufbau, an attempt to learn all knowledge from sense datum, using the notion of "remembered as similar" to bind them into clusters, which would eventually map into language.

Proponents of behaviorism argued that language may be learned through a form of operant conditioning. In B. F. Skinner's *Verbal Behavior* (1957), he suggested that the successful use of a sign, such as a word or lexical unit, given a certain stimulus, reinforces its "momentary" or contextual probability.

Since operant conditioning is contingent on reinforcement by rewards, a child would learn that a specific combination of sounds stands for a specific thing through repeated successful associations made between the two. A "successful" use of a sign would be one in which the child is understood (for example, a child saying "up" when he or she wants to be picked up) and rewarded with the desired response from another person, thereby reinforcing the child's understanding of the meaning of that word and making it more likely that he or she

will use that word in a similar situation in the future. Some empiricist theories of language acquisition include the statistical learning theory. Charles F. Hockett of language acquisition, relational frame theory, functionalist linguistics, social interactionist theory, and usage-based language acquisition.

Skinner's behaviorist idea was strongly attacked by Noam Chomsky in a review article in 1959, calling it "largely mythology" and a "serious delusion." Arguments against Skinner's idea of language acquisition through operant conditioning include the fact that children often ignore language corrections from adults. Instead, children typically follow a pattern of using an irregular form of a word correctly, making errors later on, and eventually returning to the proper use of the word. For example, a child may correctly learn the word "gave" (past tense of "give"), and later on use the word "gived". Eventually, the child will typically go back to using the correct word, "gave".

Chomsky claimed the pattern is difficult to attribute to Skinner's idea of operant conditioning as the primary way that children acquire language. Chomsky argued that if language were solely acquired through behavioral conditioning, children would not likely learn the proper use of a word and suddenly use the word incorrectly. Chomsky believed that Skinner failed to account for the central role of syntactic knowledge in language competence. Chomsky also rejected the term "learning", which Skinner used to claim that children "learn" language through operant conditioning. Instead, Chomsky argued for a mathematical approach to language acquisition, based on a study of syntax.

As a typically human phenomenon

The capacity to acquire and use language is a key aspect that distinguishes humans from other beings. Although it is difficult to pin down what aspects of language are uniquely human, there are a few design features that can be found in all known forms of human language, but that are missing from forms of animal communication. For example, many animals are able to communicate with each other by signaling to the things around them, but this kind of communication lacks the arbitrariness of human vernaculars (in that there is nothing about the sound of the word "dog" that would hint at its meaning). Other forms of animal communication may utilize arbitrary sounds, but are unable to combine those sounds in different ways to create completely novel messages that can then be automatically understood by another. Hockett called this design feature of human language "productivity". It is crucial to the understanding of human language acquisition that humans are not limited to a finite set of words, but, rather, must be able to understand and utilize a complex system that allows for an infinite number of possible messages. So, while many forms of animal communication exist, they differ from human language in that they have a limited range of vocabulary tokens, and the vocabulary items are not combined syntactically to create phrases.

Herbert S. Terrace conducted a study on a chimpanzee known as NimChimpsky in an attempt to teach him American Sign Language. This study was an attempt to further research done with a chimpanzee named Washoe, who was reportedly able to acquire American Sign Language. However, upon further

inspection, Terrace concluded that both experiments were failures. While Nim was able to acquire signs, he never acquired a knowledge of grammar, and was unable to combine signs in a meaningful way. Researchers noticed that "signs that seemed spontaneous were, in fact, cued by teachers", and not actually productive. When Terrace reviewed Project Washoe, he found similar results. He postulated that there is a fundamental difference between animals and humans in their motivation to learn language; animals, such as in Nim's case, are motivated only by physical reward, while humans learn language in order to "create a new type of communication".

In another language acquisition study, Jean-Marc-Gaspard Itard attempted to teach Victor of Aveyron, a feral child, how to speak. Victor was able to learn a few words, but ultimately never fully acquired language. Slightly more successful was a study done on Genie, another child never introduced to society. She had been entirely isolated for the first thirteen years of her life by her father. Caretakers and researchers attempted to measure her ability to learn a language. She was able to acquire a large vocabulary, but never acquired grammatical knowledge. Researchers concluded that the theory of a critical period was true; Genie was too old to learn how to speak productively, although she was still able to comprehend language.

General approaches

A major debate in understanding language acquisition is how these capacities are picked up by infants from the linguistic input. Input in the linguistic contextis defined as "All words, contexts, and other forms of language to which a learner is

exposed, relative to acquired proficiency in first or second languages". Nativists such as Chomsky have focused on the hugely complex nature of human grammars, the finiteness and ambiguity of the input that children receive, and the relatively limited cognitive abilities of an infant.

From these characteristics, they conclude that the process of language acquisition in infants must be tightly constrained and guided by the biologically given characteristics of the human brain. Otherwise, they argue, it is extremely difficult to explain how children, within the first five years of life, routinely master the complex, largely tacit grammatical rules of their native language. Additionally, the evidence of such rules in their native language is all indirect— adult speech to children cannot encompass all of what children know by the time they've acquired their native language.

Other scholars, however, have resisted the possibility that infants' routine success at acquiring the grammar of their native language requires anything more than the forms of learning seen with other cognitive skills, including such mundane motor skills as learning to ride a bike. In particular, there has been resistance to the possibility that human biology includes any form of specialization for language. This conflict is often referred to as the "nature and nurture" debate. Of course, most scholars acknowledge that certain aspects of language acquisition must result from the specific ways in which the human brain is "wired" (a "nature" component, which accounts for the failure of non-human species to acquire human languages) and that certain others are shaped by the particular language environment in which a person is raised (a "nurture" component, which accounts for the fact that humans

raised in different societies acquire different languages). The as-yet unresolved question is the extent to which the specific cognitive capacities in the "nature" component are also used outside of language.

Emergentism

Emergentist theories, such as Brian MacWhinney'scompetition model, posit that language acquisition is a cognitive process that emerges from the interaction of biological pressures and the environment. According to these theories, neither nature nor nurture alone is sufficient to trigger language learning; both of these influences must work together in order to allow children to acquire a language. The proponents of these theories argue that general cognitive processes subserve language acquisition and that the end result of these processes is language-specific phenomena, such as word learning and grammar acquisition. The findings of many empirical studies support the predictions of these theories, suggesting that language acquisition is a more complex process than many have proposed.

Empiricism

Although Chomsky's theory of a generative grammar has been enormously influential in the field of linguistics since the 1950s, many criticisms of the basic assumptions of generative theory have been put forth by cognitive-functional linguists, who argue that language structure is created through language use. These linguists argue that the concept of a language acquisition device (LAD) is unsupported by evolutionary anthropology, which tends to show a gradual adaptation of the

human brain and vocal cords to the use of language, rather than a sudden appearance of a complete set of binary parameters delineating the whole spectrum of possible grammars ever to have existed and ever to exist. On the other hand, cognitive-functional theorists use this anthropological data to show how human beings have evolved the capacity for grammar and syntax to meet our demand for linguistic symbols. (Binary parameters are common to digital computers, but may not be applicable to neurological systems such as the human brain.)

Further, the generative theory has several constructs (such as movement, empty categories, complex underlying structures, and strict binary branching) that cannot possibly be acquired from any amount of linguistic input. It is unclear that human language is actually *anything like* the generative conception of it. Since language, as imagined by nativists, is unlearnably complex, subscribers to this theory argue that it must, therefore, be innate. Nativists hypothesize that some features of syntactic categories exist even before a child is exposed to any experience - categories on which children map words of their language as they learn their native language. A different theory of language, however, may yield different conclusions. While all theories of language acquisition posit some degree of innateness, they vary in how much value they place on this innate capacity to acquire language. Empiricism places less value on the innate knowledge, arguing instead that the input, combined with both general and language-specific learning capacities, is sufficient for acquisition.

Since 1980, linguists studying children, such as Melissa Bowerman and Asifa Majid, and psychologists following Jean

Piaget, like Elizabeth Bates and Jean Mandler, came to suspect that there may indeed be many learning processes involved in the acquisition process, and that ignoring the role of learning may have been a mistake.

In recent years, the debate surrounding the nativist position has centered on whether the inborn capabilities are language-specific or domain-general, such as those that enable the infant to visually make sense of the world in terms of objects and actions.

The anti-nativist view has many strands, but a frequent theme is that language emerges from usage in social contexts, using learning mechanisms that are a part of an innate general cognitive learning apparatus. This position has been championed by David M. W. Powers, Elizabeth Bates, Catherine Snow, AnatNinio, Brian MacWhinney, Michael Tomasello, Michael Ramscar, William O'Grady, and others. Philosophers, such as Fiona Cowie and Barbara Scholz with Geoffrey Pullum have also argued against certain nativist claims in support of empiricism.

The new field of cognitive linguistics has emerged as a specific counter to Chomsky's Generative Grammar and to Nativism.

Statistical learning

Some language acquisition researchers, such as Elissa Newport, Richard Aslin, and Jenny Saffran, emphasize the possible roles of general learning mechanisms, especially statistical learning, in language acquisition. The development of connectionist models that when implemented are able to successfully learn words and syntactical conventions supports

the predictions of statistical learning theories of language acquisition, as do empirical studies of children's detection of word boundaries. In a series of connectionist model simulations, Franklin Chang has demonstrated that such a domain general statistical learning mechanism could explain a wide range of language structure acquisition phenomena.

Statistical learning theory suggests that, when learning language, a learner would use the natural statistical properties of language to deduce its structure, including sound patterns, words, and the beginnings of grammar.

That is, language learners are sensitive to how often syllable combinations or words occur in relation to other syllables. Infants between 21 and 23 months old are also able to use statistical learning to develop "lexical categories", such as an animal category, which infants might later map to newly learned words in the same category. These findings suggest that early experience listening to language is critical to vocabulary acquisition.

The statistical abilities are effective, but also limited by what qualifies as input, what is done with that input, and by the structure of the resulting output. One should also note that statistical learning (and more broadly, distributional learning) can be accepted as a component of language acquisition by researchers on either side of the "nature and nurture" debate.

From the perspective of that debate, an important question is whether statistical learning can, by itself, serve as an alternative to nativist explanations for the grammatical constraints of human language.

Chunking

The central idea of these theories is that language development occurs through the incremental acquisition of meaningful chunks of elementary constituents, which can be words, phonemes, or syllables.

Recently, this approach has been highly successful in simulating several phenomena in the acquisition of syntactic categories and the acquisition of phonological knowledge.

Chunking theories of language acquisition constitute a group of theories related to statistical learning theories, in that they assume that the input from the environment plays an essential role; however, they postulate different learning mechanisms.

Researchers at the Max Planck Institute for Evolutionary Anthropology have developed a computer model analyzing early toddler conversations to predict the structure of later conversations.

They showed that toddlers develop their own individual rules for speaking, with 'slots' into which they put certain kinds of words. A significant outcome of this research is that rules inferred from toddler speech were better predictors of subsequent speech than traditional grammars.

This approach has several features that make it unique: the models are implemented as computer programs, which enables clear-cut and quantitative predictions to be made; they learn from naturalistic input—actual child-directed utterances; and attempt to create their own utterances, the model was tested in languages including English, Spanish, and German. Chunking

for this model was shown to be most effective in learning a first language but was able to create utterances learning a second language.

Relational frame theory

The relational frame theory (RFT) (Hayes, Barnes-Holmes, Roche, 2001), provides a wholly selectionist/learning account of the origin and development of language competence and complexity. Based upon the principles of Skinnerian behaviorism, RFT posits that children acquire language purely through interacting with the environment.

RFT theorists introduced the concept of functional contextualism in language learning, which emphasizes the importance of predicting and influencing psychological events, such as thoughts, feelings, and behaviors, by focusing on manipulable variables in their own context. RFT distinguishes itself from Skinner's work by identifying and defining a particular type of operant conditioning known as derived relational responding, a learning process that, to date, appears to occur only in humans possessing a capacity for language. Empirical studies supporting the predictions of RFT suggest that children learn language through a system of inherent reinforcements, challenging the view that language acquisition is based upon innate, language-specific cognitive capacities.

Social interactionism

Social interactionist theory is an explanation of language development emphasizing the role of social interaction between the developing child and linguistically knowledgeable adults. It

is based largely on the socio-cultural theories of Soviet psychologist Lev Vygotsky, and was made prominent in the Western world by Jerome Bruner.

Unlike other approaches, it emphasizes the role of feedback and reinforcement in language acquisition. Specifically, it asserts that much of a child's linguistic growth stems from modeling of and interaction with parents and other adults, who very frequently provide instructive correction. It is thus somewhat similar to behaviorist accounts of language learning. It differs substantially, though, in that it posits the existence of a social-cognitive model and other mental structures within children (a sharp contrast to the "black box" approach of classical behaviorism).

Another key idea within the theory of social interactionism is that of the zone of proximal development. This is a theoretical construct denoting the set of tasks a child is capable of performing with guidance but not alone. As applied to language, it describes the set of linguistic tasks (for example, proper syntax, suitable vocabulary usage) that a child cannot carry out on its own at a given time, but can learn to carry out if assisted by an able adult.

Syntax, morphology, and generative grammar

As syntax began to be studied more closely in the early 20th century in relation to language learning, it became apparent to linguists, psychologists, and philosophers that knowing a language was not merely a matter of associating words with concepts, but that a critical aspect of language involves knowledge of how to put words together; sentences are usually

needed in order to communicate successfully, not just isolated words. A child will use short expressions such as *Bye-bye Mummy* or *All-gone milk*, which actually are combinations of individual nouns and an operator, before s/he begins to produce gradually more complex sentences. In the 1990s, within the principles and parameters framework, this hypothesis was extended into a maturation-based structure building model of child language regarding the acquisition of functional categories. In this model, children are seen as gradually building up more and more complex structures, with lexical categories (like noun and verb) being acquired before functional-syntactic categories (like determiner and complementiser). It is also often found that in acquiring a language, the most frequently used verbs are irregular verbs. In learning English, for example, young children first begin to learn the past tense of verbs individually. However, when they acquire a "rule", such as adding *-ed* to form the past tense, they begin to exhibit occasional overgeneralization errors (e.g. "runned", "hitted") alongside correct past tense forms. One influential proposal regarding the origin of this type of error suggests that the adult state of grammar stores each irregular verb form in memory and also includes a "block" on the use of the regular rule for forming that type of verb. In the developing child's mind, retrieval of that "block" may fail, causing the child to erroneously apply the regular rule instead of retrieving the irregular.

A Merge (linguistics)-based Theory

In Bare-Phrase structure (Minimalist Program), since theory-internal considerations define the specifier position of an internal-merge projection (phases vP and CP) as the only type

of host which could serve as potential landing-sites for move-based elements displaced from lower down within the base-generated VP structure – e.g., A-movement such as passives (["The apple was eaten by [John (ate the apple)"]]), or raising ["Some work does seem to remain [(There) does seem to remain (some work)"]])—as a consequence, any strong version of a Structure building model of child language which calls for an exclusive "external-merge/argument structure stage" prior to an "internal-merge/scope-discourse related stage" would claim that young children's stage-1 utterances lack the ability to generate and host elements derived via movement operations. In terms of a Merge-based theory of language acquisition, complements and specifiers are simply notations for first-merge (= "complement-of" [head-complement]), and later second-merge (= "specifier-of" [specifier-head], with merge always forming to a head. First-merge establishes only a set {a, b} and is not an ordered pair—e.g., an {N, N}-compound of 'boat-house' would allow the ambiguous readings of either 'a kind of house' and/or 'a kind of boat'.

It is only with second-merge that order is derived out of a set {a {a, b}} which yields the recursive properties of syntax—e.g., a 'house-boat' {house {house, boat}} now reads unambiguously only as a 'kind of boat'. It is this property of recursion that allows for projection and labeling of a phrase to take place; in this case, that the Noun 'boat' is the Head of the compound, and 'house' acting as a kind of specifier/modifier. External-merge (first-merge) establishes substantive 'base structure' inherent to the VP, yielding theta/argument structure, and may go beyond the lexical-category VP to involve the functional-category light verb vP. Internal-merge (second-merge) establishes more formal aspects related to edge-

properties of scope and discourse-related material pegged to CP. In a Phase-based theory, this twin vP/CP distinction follows the "duality of semantics" discussed within the Minimalist Program, and is further developed into a dual distinction regarding a probe-goal relation. As a consequence, at the "external/first-merge-only" stage, young children would show an inability to interpret readings from a given ordered pair, since they would only have access to the mental parsing of a non-recursive set. (See Roeper for a full discussion of recursion in child language acquisition). In addition to word-order violations, other more ubiquitous results of a first-merge stage would show that children's initial utterances lack the recursive properties of inflectional morphology, yielding a strict Non-inflectional stage-1, consistent with an incremental Structure-building model of child language.

Generative grammar, associated especially with the work of Noam Chomsky, is currently one of the approaches to explaining children's acquisition of syntax. Its leading idea is that human biology imposes narrow constraints on the child's "hypothesis space" during language acquisition. In the principles and parameters framework, which has dominated generative syntax since Chomsky's (1980) *Lectures on Government and Binding: The Pisa Lectures*, the acquisition of syntax resembles ordering from a menu: the human brain comes equipped with a limited set of choices from which the child selects the correct options by imitating the parents' speech while making use of the context.

An important argument which favors the generative approach, is the poverty of the stimulus argument. The child's input (a finite number of sentences encountered by the child, together

with information about the context in which they were uttered) is, in principle, compatible with an infinite number of conceivable grammars. Moreover, rarely can children rely on corrective feedback from adults when they make a grammatical error; adults generally respond and provide feedback regardless of whether a child's utterance was grammatical or not, and children have no way of discerning if a feedback response was intended to be a correction. Additionally, when children do understand that they are being corrected, they don't always reproduce accurate restatements. Yet, barring situations of medical abnormality or extreme privation, all children in a given speech-community converge on very much the same grammar by the age of about five years. An especially dramatic example is provided by children who, for medical reasons, are unable to produce speech and, therefore, can never be corrected for a grammatical error but nonetheless, converge on the same grammar as their typically-developing peers, according to comprehension-based tests of grammar.

Considerations such as those have led Chomsky, Jerry Fodor, Eric Lenneberg and others to argue that the types of grammar the child needs to consider must be narrowly constrained by human biology (the nativist position). These innate constraints are sometimes referred to as universal grammar, the human "language faculty", or the "language instinct".

Representation in the brain

Recent advances in functional neuroimaging technology have allowed for a better understanding of how language acquisition is manifested physically in the brain. Language acquisition almost always occurs in children during a period of rapid

increase in brain volume. At this point in development, a child has many more neural connections than he or she will have as an adult, allowing for the child to be more able to learn new things than he or she would be as an adult.

Sensitive period

Language acquisition has been studied from the perspective of developmental psychology and neuroscience, which looks at learning to use and understand language parallel to a child's brain development. It has been determined, through empirical research on developmentally normal children, as well as through some extreme cases of language deprivation, that there is a "sensitive period" of language acquisition in which human infants have the ability to learn any language. Several researchers have found that from birth until the age of six months, infants can discriminate the phonetic contrasts of all languages.

Researchers believe that this gives infants the ability to acquire the language spoken around them. After this age, the child is able to perceive only the phonemes specific to the language being learned. The reduced phonemic sensitivity enables children to build phonemic categories and recognize stress patterns and sound combinations specific to the language they are acquiring. As Wilder Penfield noted, "Before the child begins to speak and to perceive, the uncommitted cortex is a blank slate on which nothing has been written. In the ensuing years much is written, and the writing is normally never erased. After the age of ten or twelve, the general functional connections have been established and fixed for the speech cortex." According to the sensitive or critical period

models, the age at which a child acquires the ability to use language is a predictor of how well he or she is ultimately able to use language. However, there may be an age at which becoming a fluent and natural user of a language is no longer possible; Penfield and Roberts (1959) cap their sensitive period at nine years old. The human brain may be automatically wired to learn languages, but this ability does not last into adulthood in the same way that it exists during childhood. By around age 12, language acquisition has typically been solidified, and it becomes more difficult to learn a language in the same way a native speaker would. Just like children who speak, deaf children go through a critical period for learning language. Deaf children who acquire their first language later in life show lower performance in complex aspects of grammar. At that point, it is usually a second language that a person is trying to acquire and not a first.

Assuming that children are exposed to language during the critical period, acquiring language is almost never missed by cognitively normal children. Humans are so well-prepared to learn language that it becomes almost impossible not to. Researchers are unable to experimentally test the effects of the sensitive period of development on language acquisition, because it would be unethical to deprive children of language until this period is over. However, case studies on abused, language-deprivedchildren show that they exhibit extreme limitations in language skills, even after instruction.

At a very young age, children can distinguish different sounds but cannot yet produce them. During infancy, children begin to babble. Deaf babies babble in the same patterns as hearing babies do, showing that babbling is not a result of babies

simply imitating certain sounds, but is actually a natural part of the process of language development. Deaf babies do, however, often babble less than hearing babies, and they begin to babble later on in infancy—at approximately 11 months as compared to approximately 6 months for hearing babies.

Prelinguistic language abilities that are crucial for language acquisition have been seen even earlier than infancy. There have been many different studies examining different modes of language acquisition prior to birth. The study of language acquisition in fetuses began in the late 1980s when several researchers independently discovered that very young infants could discriminate their native language from other languages. In *Mehler et al. (1988)*, infants underwent discrimination tests, and it was shown that infants as young as 4 days old could discriminate utterances in their native language from those in an unfamiliar language, but could not discriminate between two languages when neither was native to them. These results suggest that there are mechanisms for fetal auditory learning, and other researchers have found further behavioral evidence to support this notion. Fetus auditory learning through environmental habituation has been seen in a variety of different modes, such as fetus learning of familiar melodies (Hepper, 1988), story fragments (DeCasper& Spence, 1986), recognition of mother's voice (Kisilevsky, 2003), and other studies showing evidence of fetal adaptation to native linguistic environments (Moon, Cooper & Fifer, 1993).

Prosody is the property of speech that conveys an emotional state of the utterance, as well as the intended form of speech, for example, question, statement or command. Some researchers in the field of developmental neuroscience argue

that fetal auditory learning mechanisms result solely from discrimination of prosodic elements. Although this would hold merit in an evolutionary psychology perspective (i.e. recognition of mother's voice/familiar group language from emotionally valent stimuli), some theorists argue that there is more than prosodic recognition in elements of fetal learning. Newer evidence shows that fetuses not only react to the native language differently from non-native languages, but that fetuses react differently and can accurately discriminate between native and non-native vowel sounds (Moon, Lagercrantz, &Kuhl, 2013). Furthermore, a 2016 study showed that newborn infants encode the edges of multisyllabic sequences better than the internal components of the sequence (Ferry et al., 2016).

Together, these results suggest that newborn infants have learned important properties of syntactic processing in utero, as demonstrated by infant knowledge of native language vowels and the sequencing of heard multisyllabic phrases. This ability to sequence specific vowels gives newborn infants some of the fundamental mechanisms needed in order to learn the complex organization of a language. From a neuroscientific perspective, neural correlates have been found that demonstrate human fetal learning of speech-like auditory stimuli that most other studies have been analyzing (Partanen et al., 2013). In a study conducted by Partanen et al. (2013), researchers presented fetuses with certain word variants and observed that these fetuses exhibited higher brain activity in response to certain word variants as compared to controls. In this same study, "a significant correlation existed between the amount of prenatal exposure and brain activity, with greater activity being associated with a higher amount of prenatal speech exposure,"

pointing to the important learning mechanisms present before birth that are fine-tuned to features in speech (Partanen et al., 2013).

Vocabulary acquisition

The capacity to acquire the ability to incorporate the pronunciation of new words depends upon many factors. First, the learner needs to be able to hear what they are attempting to pronounce. Also required is the capacity to engage in speech repetition. Children with reduced ability to repeat non-words (a marker of speech repetition abilities) show a slower rate of vocabulary expansion than children with normal ability. Several computational models of vocabulary acquisition have been proposed.

Various studies have shown that the size of a child's vocabulary by the age of 24 months correlates with the child's future development and language skills. A lack of language richness by this age has detrimental and long-term effects on the child's cognitive development, which is why it is so important for parents to engage their infants in language. If a child knows fifty or fewer words by the age of 24 months, he or she is classified as a late-talker, and future language development, like vocabulary expansion and the organization of grammar, is likely to be slower and stunted.

Two more crucial elements of vocabulary acquisition are word segmentation and statistical learning (described above). Word segmentation, or the ability to break down words into syllables from fluent speech can be accomplished by eight-month-old

infants. By the time infants are 17 months old, they are able to link meaning to segmented words.

Recent evidence also suggests that motor skills and experiences may influence vocabulary acquisition during infancy. Specifically, learning to sit independently between 3 and 5 months of age has been found to predict receptive vocabulary at both 10 and 14 months of age, and independent walking skills have been found to correlate with language skills at around 10 to 14 months of age.

These findings show that language acquisition is an embodied process that is influenced by a child's overall motor abilities and development. Studies have also shown a correlation between socioeconomic status and vocabulary acquisition.

Meaning

Children learn, on average, ten to fifteen new word meanings each day, but only one of these can be accounted for by direct instruction. The other nine to fourteen word meanings must have been acquired in some other way. It has been proposed that children acquire these meanings through processes modeled by latent semantic analysis; that is, when they encounter an unfamiliar word, children use contextual information to guess its rough meaning correctly.

A child may expand the meaning and use of certain words that are already part of its mental lexicon in order to denominate anything that is somehow related but for which it does not know the specific word. For instance, a child may broaden the use of *mummy* and *dada* in order to indicate anything that

belongs to its mother or father, or perhaps every person who resembles its own parents; another example might be to say *rain* while meaning *I don't want to go out*.

There is also reason to believe that children use various heuristics to infer the meaning of words properly. Markman and others have proposed that children assume words to refer to objects with similar properties ("cow" and "pig" might both be "animals") rather than to objects that are thematically related ("cow" and "milk" are probably not both "animals"). Children also seem to adhere to the "whole object assumption" and think that a novel label refers to an entire entity rather than to one of its parts. This assumption along with other resources, such as grammar and morphological cues or lexical constraints, may help aid the child in acquiring word meaning, but conclusions based on such resources may sometimes conflict.

Genetic and neurocognitive research

According to several linguists, neurocognitive research has confirmed many standards of language learning, such as: "learning engages the entire person (cognitive, affective, and psychomotor domains), the human brain seeks patterns in its searching for meaning, emotions affect all aspects of learning, retention and recall, past experience always affects new learning, the brain's working memory has a limited capacity, lecture usually results in the lowest degree of retention, rehearsal is essential for retention, practice [alone] does not make perfect, and each brain is unique" (Sousa, 2006, p. 274). In terms of genetics, the gene ROBO1 has been associated with phonological buffer integrity or length.

Genetic research has found two major factors predicting successful language acquisition and maintenance. These include inherited intelligence, and the lack of genetic anomalies that may cause speech pathologies, such as mutations in the FOXP2 gene which cause verbal dyspraxia. The role of inherited intelligence increases with age, accounting for 20% of IQ variation in infants, and for 60 % in adults. It affects a vast variety of language-related abilities, from spatio-motor skills to writing fluency. There have been debates in linguistics, philosophy, psychology, and genetics, with some scholars arguing that language is fully or mostly innate, but the research evidence points to genetic factors only working in interaction with environmental ones.

Although it is difficult to determine without invasive measures which exact parts of the brain become most active and important for language acquisition, fMRI and PET technology has allowed for some conclusions to be made about where language may be centered. Kuniyoshi Sakai has proposed, based on several neuroimaging studies, that there may be a "grammar center" in the brain, whereby language is primarily processed in the left lateral premotor cortex (located near the pre central sulcus and the inferior frontal sulcus). Additionally, these studies have suggested that first language and second language acquisition may be represented differently in the cortex. In a study conducted by Newman et al., the relationship between cognitive neuroscience and language acquisition was compared through a standardized procedure involving native speakers of English and native Spanish speakers who all had a similar length of exposure to the English language (averaging about 26 years). It was concluded that the brain does in fact process languages

differently, but rather than being related to proficiency levels, language processing relates more to the function of the brain itself.

During early infancy, language processing seems to occur over many areas in the brain. However, over time, it gradually becomes concentrated into two areas – Broca's area and Wernicke's area. Broca's area is in the left frontal cortex and is primarily involved in the production of the patterns in vocal and sign language. Wernicke's area is in the left temporal cortex and is primarily involved in language comprehension. The specialization of these language centers is so extensive that damage to them can result in aphasia.

Artificial intelligence

Some algorithms for language acquisition are based on statistical machine translation. Language acquisition can be modeled as a machine learning process, which may be based on learning semantic parsers or grammar induction algorithms.

Prelingual deafness

Prelingual deafness is defined as hearing loss that occurred at birth or before an individual has learned to speak. In the United States, 2 to 3 out of every 1000 children are born deaf or hard of hearing. Even though it might be presumed that deaf children acquire language in different ways since they are not receiving the same auditory input as hearing children, many research findings indicate that deaf children acquire language in the same way that hearing children do and when

given the proper language input, understand and express language just as well as their hearing peers. Babies who learn sign language produce signs or gestures that are more regular and more frequent than hearing babies acquiring spoken language. Just as hearing babies babble, deaf babies acquiring sign language will babble with their hands, otherwise known as manual babbling. Therefore, as many studies have shown, language acquisition by deaf children parallel the language acquisition of a spoken language by hearing children because humans are biologically equipped for language regardless of the modality.

Signed language acquisition

Deaf children's visual-manual language acquisition not only parallel spoken language acquisition but by the age of 30 months, most deaf children that were exposed to a visual language had a more advanced grasp with subject-pronoun copy rules than hearing children. Their vocabulary bank at the ages of 12–17 months exceed that of a hearing child's, though it does even out when they reach the two-word stage. The use of space for absent referents and the more complex handshapes in some signs prove to be difficult for children between 5 and 9 years of age because of motor development and the complexity of remembering the spatial use.

Cochlear implants

Other options besides sign language for kids with prelingual deafness include the use of hearing aids to strengthen remaining sensory cells or cochlear implants to stimulate the hearing nerve directly. Cochlear Implants are hearing devices

that are placed behind the ear and contain a receiver and electrodes which are placed under the skin and inside the cochlea. Despite these developments, there is still a risk that prelingually deaf children may not develop good speech and speech reception skills. Although cochlear implants produce sounds, they are unlike typical hearing and deaf and hard of hearing people must undergo intensive therapy in order to learn how to interpret these sounds. They must also learn how to speak given the range of hearing they may or may not have. However, deaf children of deaf parents tend to do better with language, even though they are isolated from sound and speech because their language uses a different mode of communication that is accessible to them; the visual modality of language.

Although cochlear implants were initially approved for adults, now there is pressure to implant children early in order to maximize auditory skills for mainstream learning which in turn has created controversy around the topic. Due to recent advances in technology, cochlear implants allow some deaf people to acquire some sense of hearing. There are interior and exposed exterior components that are surgically implanted. Those who receive cochlear implants earlier on in life show more improvement on speech comprehension and language. Spoken language development does vary widely for those with cochlear implants though due to a number of different factors including: age at implantation, frequency, quality and type of speech training. Some evidence suggests that speech processing occurs at a more rapid pace in some prelingually deaf children with cochlear implants than those with traditional hearing aids. However, cochlear implants may not always work.

Research shows that people develop better language with a cochlear implant when they have a solid first language to rely on to understand the second language they would be learning. In the case of prelingually deaf children with cochlear implants, a signed language, like American Sign Language would be an accessible language for them to learn to help support the use of the cochlear implant as they learn a spoken language as their L2. Without a solid, accessible first language, these children run the risk of language deprivation, especially in the case that a cochlear implant fails to work. They would have no access to sound, meaning no access to the spoken language they are supposed to be learning. If a signed language was not a strong language for them to use and neither was a spoken language, they now have no access to any language and run the risk of missing their critical period.

Sentence processing

Sentence processing takes place whenever a reader or listener processes a language utterance, either in isolation or in the context of a conversation or a text. Many studies of the human language comprehension process have focused on reading of single utterances (sentences) without context. Extensive research has shown that language comprehension is affected by context preceding a given utterance as well as many other factors.

Ambiguity

Sentence comprehension has to deal with ambiguity in spoken and written utterances, for example lexical, structural, and

semantic ambiguities. Ambiguity is ubiquitous, but people usually resolve it so effortlessly that they do not even notice it. For example, the sentence *Time flies like an arrow* has (at least) the interpretations *Time moves as quickly as an arrow*, *A special kind of fly, called time fly, likes arrows* and *Measure the speed of flies like you would measure the speed of an arrow*. Usually, readers will be aware of only the first interpretation. Educated readers though, spontaneously think about the arrow of time but inhibit that interpretation because it deviates from the original phrase and the temporal lobe acts as a switch.

Instances of ambiguity can be classified as **local** or **global** ambiguities. A sentence is globally ambiguous if it has two distinct interpretations.

Examples are sentences like *Someone shot the servant of the actress who was on the balcony* (was it the servant or the actress who was on the balcony?) or *The cop chased the criminal with a fast car* (did the cop or the criminal have a fast car?). Comprehenders may have a preferential interpretation for either of these cases, but syntactically and semantically, neither of the possible interpretations can be ruled out.

Local ambiguities persist only for a short amount of time as an utterance is heard or written and are resolved during the course of the utterance so the complete utterance has only one interpretation. Examples include sentences like*The critic wrote the book was enlightening*, which is ambiguous when *The critic wrote the book* has been encountered, but *was enlightening* remains to be processed. Then, the sentence could end, stating that the critic is the author of the book, or it could go on to clarify that the critic wrote something about a book. The

ambiguity ends at *was enlightening*, which determines that the second alternative is correct.

When readers process a local ambiguity, they settle on one of the possible interpretations immediately without waiting to hear or read more words that might help decide which interpretation is correct (the behaviouris called*incremental processing*). If readers are surprised by the turn the sentence really takes, processing is slowed and is visible for example in reading times. Locally-ambiguous sentences have, therefore, been used as test cases to investigate the influence of a number of different factors on human sentence processing. If a factor helps readers to avoid difficulty, it is clear that the factor plays a factor in sentence processing.

Theories

Experimental research has spawned a large number of hypotheses about the architecture and mechanisms of sentence comprehension. Issues like modularity versus interactive processing and serial versus parallel computation of analyses have been theoretical divides in the field.

Architectural issues

Modular vs. interactive

A modular view of sentence processing assumes that each factor involved in sentence processing is computed in its own module, which has limited means of communication with the other modules. For example, syntactic analysis creation takes place without input from semantic analysis or context-

dependent information, which are processed separately. A common assumption of modular accounts is a *feed-forward* architecture in which the output of one processing step is passed on to the next step without feedback mechanisms that would allow the output of the first module to be corrected. Syntactic processing is usually taken to be the most basic analysis step, which feeds into semantic processing and the inclusion of other information.

A separate mental module parses sentences and lexical access happens first. Then, one syntactic hypothesis is considered at a time. There is no initial influence of meaning, or semantic. Sentence processing is supported by a temporo-frontal network. Within the network, temporal regions subserve aspects of identification and frontal regions the building of syntactic and semantic relations. Temporal analyses of brain activation within this network support syntax-first models because they reveal that building of syntactic structure precedes semantic processes and that these interact only during a later stage.

Interactive accounts assume that all available information is processed at the same time and can immediately influence the computation of the final analysis. In the interactive model of sentence processing, there is no separate module for parsing. Lexical access, syntactic structure assignment, and meaning assignment happen at the same time in parallel. Several syntactic hypotheses can be considered at a time. The interactive model demonstrates an on-line interaction between the structural and lexical and phonetic levels of sentence processing. Each word, as it is heard in the context of normal discourse, is immediately entered into the processing system at

all levels of description, and is simultaneously analyzed at all these levels in the light of whatever information is available at each level at that point in the processing of the sentence. Interactive models of language processing assume that information flows both bottom-up and top-down, so that the representations formed at each level may be influenced by higher as well as lower levels.

A framework called the interactive activation framework that embeds this key assumption among others, including the assumption that influences from different sources are combined nonlinearly. The nonlinearity means that information that may be decisive under some circumstances may have little or no effect under other conditions. In the interactive activation framework, the knowledge that guides processing is stored in the connections between units on the same and adjacent levels. The processing units that they connect may receive input from a number of different sources, which allows the knowledge that guides processing to be completely local while, at the same time, allowing the results of processing at one level to influence processing at other levels, both above and below. A basic assumption of the framework is that processing interactions are always reciprocal; it is this bi-directional characteristic that makes the system interactive. Bi-directional excitatory interactions between levels allow mutual simultaneous constraint among adjacent levels, and bi-directional inhibitory interactions within a level allow for competition among mutually incompatible interpretations of a portion of an input. The between-level excitatory interactions are captured in the models in two-way excitatory connections between mutually compatible processing units. Syntactic ambiguities are in fact based at the lexical level. In addition,

more recent studies with more sensitive eye tracking machines have shown early context effects. Frequency and contextual information will modulate the activation of alternatives even when they are resolved in favor of the simple interpretation. Structural simplicity is cofounded with frequency, which goes against the garden path theory

Serial vs. parallel

Serial accounts assume that humans construct only one of the possible interpretations at first and try another only if the first one turns out to be wrong. Parallel accounts assume the construction of multiple interpretations at the same time. To explain why comprehenders are usually only aware of one possible analysis of what they hear, models can assume that all analyses ranked, and the highest-ranking one is entertained.

Models

There are a number of influential models of human sentence processing that draw on different combinations of architectural choices.

Garden path model

The garden path model (Frazier 1987) is a serial modular parsing model. It proposes that a single parse is constructed by a syntactic module. Contextual and semantic factors influence processing at a later stage and can induce re-analysis of the syntactic parse. Re-analysis is costly and leads to an observable slowdown in reading. When the parser

encounters an ambiguity, it is guided by two principles: late closure and minimal attachment. The model has been supported with research on the early left anterior negativity, an event-related potential often elicited as a response to phrase structure violations.

Late closure causes new words or phrases to be attached to the current clause. For example, "John said he would leave yesterday" would be parsed as *John said (he would leave yesterday)*, and not as *John said (he would leave) yesterday* (i.e., he spoke yesterday).

Minimal attachment is a strategy of parsimony: The parser builds the simplest syntactic structure possible (that is, the one with the fewest phrasal nodes).

Constraint-based model

Constraint-based theories of language comprehension emphasize how people make use of the vast amount of probabilistic information available in the linguistic signal. Through statistical learning, the frequencies and distribution of events in linguistic environments can be picked upon, which inform language comprehension. As such, language users are said to arrive at a particular interpretation over another during the comprehension of an ambiguous sentence by rapidly integrating these probabilistic constraints.

Good enough theory

The good enough approach to language comprehension, developed by Fernanda Ferreira and others, assumes that

listeners do not always engage in full detailed processing of linguistic input. Rather, the system has a tendency to develop shallow and superficial representations when confronted with some difficulty.

The theory takes an approach that somewhat combines both the garden path model and the constraint based model. The theory focuses on two main issues. The first is that representations formed from complex or difficult material are often shallow and incomplete. The second is that limited information sources are often consulted in cases where the comprehension system encounters difficulty. The theory can be put to test using various experiments in psycholinguistics that involve garden path misinterpretation, etc.

Methods

Behavioral tasks

In behavioral studies, subjects are often presented with linguistic stimuli and asked to perform an action. For example, they may be asked to make a judgment about a word (lexical decision), reproduce the stimulus, or name a visually presented word aloud. Speed (often reaction time: time taken to respond to the stimulus) and accuracy (proportion of correct responses) are commonly employed measures of performance in behavioral tasks. Researchers infer that the nature of the underlying process(es) required by the task gives rise to differences; slower rates and lower accuracy on these tasks are taken as measures of increased difficulty. An important component of any behavioral task is that it stays relatively

true to 'normal' language comprehension—the ability to generalize the results of any task is restricted when the task has little in common with how people actually encounter language.

A common behavioral paradigm involves priming effects, wherein participants are presented first with a prime and then with a target word. The response time for the target word is affected by the relationship between the prime and the target. For example, Fischler (1977) investigated word encoding using the lexical decision task.

She asked participants to make decisions about whether two strings of letters were English words. Sometimes the strings would be actual English words requiring a "yes" response, and other times they would be nonwordsrequiring a "no" response. A subset of the licit words were related semantically (e.g., cat-dog) while others were unrelated (e.g., bread-stem). Fischler found that related word pairs were responded to faster when compared to unrelated word pairs, which suggests that semantic relatedness can facilitate word encoding.

Eye-movements

Eye trackinghas been used to study online language processing. This method has been influential in informing knowledge of reading. Additionally, Tanenhaus et al. (1995) established the visual world paradigm, which takes advantage of eye movements to study online spoken language processing. This area of research capitalizes on the linking hypothesis that eye movements are closely linked to the current focus of attention.

Neuroimaging and evoked potentials

The rise of non-invasive techniques provides myriad opportunities for examining the brain bases of language comprehension. Common examples include positron emission tomography (PET), functional magnetic resonance imaging (fMRI), event-related potentials (ERPs) in electroencephalography (EEG) and magnetoencephalography (MEG), and transcranial magnetic stimulation (TMS).

These techniques vary in their spatial and temporal resolutions (fMRI has a resolution of a few thousand neurons per pixel, and ERP has millisecond accuracy), and each type of methodology presents a set of advantages and disadvantages for studying a particular problem in language comprehension.

Computational modeling

Computational modeling is another means by which to explore language comprehension. Models, such as those instantiated in neural networks, are particularly useful because they requires theorists to be explicit in their hypotheses and because they can be used to generate accurate predictions for theoretical models that are so complex that they render discursive analysis unreliable. A classic example of computational modeling in language research is McClelland and Elman'sTRACE model of speech perception. A model of sentence processing can be found in Hale (2011)'s 'rational' Generalized Left Corner parser. This model derives garden path effects as well as local coherence phenomena. Computational modeling can also help to relate sentence processing to other functions of language. For example, one model of ERP effects

in sentence processing (e.g., N400 and P600) argues that these phenomena arise out learning processes that support language acquisition and linguistic adaptation.

Language production

Language production is the production of spoken or written language. In psycholinguistics, it describes all of the stages between having a concept to express and translating that concept into linguistic forms. These stages have been described in two types of processing models: the lexical access models and the serial models. Through these models, psycholinguists can look into how speeches are produced in different ways, such as when the speaker is bilingual. Psycholinguists learn more about these models and different kinds of speech by using language production research methods that include collecting speech errors and elicited production tasks.

Stages involved

Language production consists of several interdependent processes which transform a nonlinguistic message into a spoken, signed, or written linguistic signal. Though the following steps proceed in this approximate order, there are plenty of interaction and communication between them. The process of message planning is an active area of psycholinguistic research, but researchers have found that it is an ongoing process throughout language production. Research suggests that messages are planned in roughly the same order that they are in an utterance. But, there is also evidence that suggests the verbs that give case may be planned earlier than

objects, even when the object is said first. After identifying a message, or part of a message, to be linguistically encoded, a speaker must selects the individual words—also known as lexical items—to represent that message. This process is called lexical selection. The words are selected based on their meaning, which in linguistics is called semantic information. Lexical selection activates the word's lemma, which contains both semantic and grammatical information about the word.

This grammatical information is then used in the next step of language production, grammatical encoding. Critical grammatical information includes characteristics such as the word's syntactic category (noun, verb, etc.), what objects it takes, and grammatical gender if it is presents in the language. Using some of these characteristics as well as information about the thematic roles of each word in the intended message, each word is then assigned the grammatical and thematic role it will have in the sentence.

Function morphemes, like the plural /s/ or the past tense /ɪd/, are added in this stage as well. After an utterance, or part of one, has been formed, it then goes through phonological encoding. In this stage of language production, the mental representation of the words to be spoken is transformed into a sequence of speech sounds to be pronounced. The speech sounds are assembled in the order they are to be produced.

The basic loop occurring in the creation of language consists of the following stages:

- Intended message
- Encode message into linguistic form

- Encode linguistic form into speech motor system
- Sound goes from speaker's mouth to hearer's ear auditory system
- Speech is decoded into linguistic form
- Linguistic form is decoded into meaning

According to the lexical access model (see section below), in terms of lexical access, two different stages of cognition are employed; thus, this concept is known as the two-stage theory of lexical access.

The first stage, lexical selection provides information about lexical items required to construct the functional level representation.

These items are retrieved according to their specific semantic and syntactic properties, but phonological forms are not yet made available at this stage. The second stage, retrieval of wordforms, provides information required for building the positional level representation.

Models

Serial model

A serial model of language production divides the process into several stages. For example, there may be one stage for determining pronunciation and a stage for determining lexical content. The serial model does not allow overlap of these stages, so they may only be completed one at a time.

Connectionist model

Several researchers have proposed a connectionist model, one notable example being Dell. According to his connectionist model, there are four layers of processing and understanding: semantic, syntactic, morphological, and phonological.

These work in parallel and in series, with activation at each level. Interference and misactivation can occur at any of these stages. Production begins with concepts, and continues down from there.

One might start with the concept of a cat: a four-legged, furry, domesticated mammal with whiskers, etc. This conceptual set would attempt to find the corresponding word {cat}. This selected word would then selects morphological and phonological data /k / at/.

The distinction of this model is that, during this process, other elements would also be primed ({rat} might be somewhat primed, for example), as they are physically similar, and so can cause conceptual interference.

Errors might also occur at the phoneme level, as many words are phonetically similar, e.g. mat. Substitutions of similar consonant sounds are more likely to occur, e.g. between plosive stop consonants such as d, p and b. Lower primed words are less likely to be chosen, but interference is thought to occur in cases of early selection, where the level of activation of the target and interference words is at the same level.

Lexical access model

This model states that the sentence is made by a sequence of processes generating differing levels of representations. For instance, the functional level representation is made on the a preverbal representation, which is essentially what the speaker seeks to express.

This level is responsible for encoding the meanings of lexical items and the way that grammar forms relationships between them. Next, the positional level representation is built, which functions to encode the phonological forms of words and the order they are found in sentence structures.

Lexical access, according to this model, is a process that encompasses two serially ordered and independent stages.

Additional aspects

Fluency

Fluency can be defined in part by prosody, which is shown graphically by a smooth intonation contour, and by a number of other elements: control of speech rate, relative timing of stressed and unstressed syllables, changes in amplitude, changes in fundamental frequency.

In other words, fluency can be described as whether someone speaks smoothly and easily. This term is used in speech-language pathology when describing disorders with stuttering or other disfluencies.

Multilingualism

Whether or not a speaker is fluent in one or more languages, the process for producing language remains the same. However, bilinguals speaking two languages within a conversation may have access to both languages at the same time. Three of the most commonly discussed models for multilingual language access are the Bilingual Interactive Activation Plus model, the Revised Hierarchical Model, and the Language Mode model:

- **Bilingual Interactive Activation Plus**, updated from a model made by Dijkstra and Van Heuven, uses solely bottom-up processing to facilitate bilingual language access. This model suggests that the lexicon for bilingual speakers combines the languages, and access occurs across both languages at the same time.
- **Revised Hierarchical Model**, developed by Kroll and Stewart, is a model suggesting that bilingual brains store meanings in a common place, word-forms are separated by language.
- **Language Mode Model**, made by Grosjean, uses two assumptions to map bilingual language production in a modular way. These assumptions are that a base language is activated in conversation, and that the speaker's other language is activated to relative degrees depending on context. De Bot describes it as overly simple for the complexity of the process and suggests it has room for expansion.

Speakers fluent in multiple languages may inhibit access to one of their languages, but this suppression can only be done once the speaker is at a certain level of proficiency in that language.

A speaker can decide to inhibit a language based on non-linguistic cues in their conversation, such as a speaker of both English and French inhibiting their French when conversing with people who only speak English. When especially proficient multilingual speakers communicate, they can participate in code-switching. Code-switching has been shown to indicate bilingual proficiency in a speaker, though it had previously been seen as a sign of poor language ability.

Research methods

There are three main types of research into language production: speech error collection, picture-naming, and elicited production. Speech error collection focuses on using the analysis of speech errors made in naturally produced speech. On the other hand, elicited production focuses on elicited speech and is conducted in a lab. Also conducted in a lab, picture-naming focuses on reaction-time data from picture-naming latencies. Although originally disparate, these three methodologies are generally looking at the same underlying processes of speech production.

Speech errors

Speech errors have been found to be common in naturally produced speech. Analysis of speech errors has found that not all are random, but rather systematic and fall into several

categories. These speech errors can demonstrate parts of the language processing system, and what happens when that system doesn't work as it should. Language production occurs quickly with speakers saying a little more than 2 words per second; so though errors occur only once out of 1,000 words, they occur relatively often throughout a speaker's day at once every 7 minutes. Some examples of these speech errors that would be collected by psycholinguists are:

- **Anticipation**: The word adds a sound from a word planned for later in the utterance.
- *target*: paddle tennis
- *produced*: taddle tennis
- **Preservation**: The word retains characteristics of a word said previously in an utterance.
- *target*: red wagon
- *produced*: red ragon
- **Blending**: More than one word is being considered in the lexicon and the two intended items "blend" into a single item.
- *target*: shout/yell
- produced: shell
- **Addition**: Additional of linguistics material added to the word.
- *target*: impossible
- *produced*: implossible
- **Substitution**: A whole word of related meaning is replacing another.
- *target*: at low speed it's too heavy
- *produced*: at low speed it's too light
- **Malapropism**: A lay term, in reference to a character Mrs. Malaprop from Sheridan's The

Rivals, referring to the incorrect substitution of words.

- Makes no delusions to the past.
- The pineapple of perfection.
- I have interceded another letter from the fellow.
- **Spoonerism**: The switching of the letters from two words in the utterance.
- *target*: slips of the tongue
- *produced*: tips of the slung

Picture-naming

Picture-naming tasks ask participants to look at pictures and name them in a certain way.

By looking at the time course for the responses in these tasks, psycholinguists can learn more about the planning involved in specific phrases. These types of tasks can be helpful for investigating cross-linguistic language production and planning processes.

Elicited Production

Elicited production tasks ask participants to respond to questions or prompts in a particular way. One of the more common types of elicited production tasks is the sentence completion task.

These tasks give the participants the beginning of a target sentence, which the participants are then asked to complete. Analyzing these completions can allow psycholinguistics to investigate errors that might be difficult to elicit otherwise.

Second-language acquisition

Second-language acquisition (SLA), sometimes called second-language learning — otherwise referred to as L2 (language 2) acquisition, is the process by which people learn a second language. Second-language acquisition is also the scientific discipline devoted to studying that process. The field of second-language acquisition is a sub-discipline of applied linguistics but also receives research attention from a variety of other disciplines, such as psychology and education.

A central theme in SLA research is that of *interlanguage:* the idea that the language that learners use is not simply the result of differences between the languages that they already know and the language that they are learning, but a complete language system in its own right, with its own systematic rules. This interlanguage gradually develops as learners are exposed to the targeted language. The order in which learners acquire features of their new language stays remarkably constant, even for learners with different native languages and regardless of whether they have had language instruction. However, languages that learners already know can have a significant influence on the process of learning a new one. This influence is known as *language transfer*.

The primary factor driving SLA appears to be the language input that learners receive. Learners become more advanced the longer they are immersed in the language they are learning and the more time they spend voluntarily reading. The input hypothesis developed by linguist Stephen Krashen theorizes that comprehensible input alone is necessary for second language acquisition. Krashen makes a distinction between

language acquisition and language learning (the acquisition–learning distinction), claiming that acquisition is a subconscious process, whereas learning is a conscious one. According to this hypothesis, the acquisition process in L2 (Language 2) is the same as L1 (Language 1) acquisition. Learning, on the other hand, refers to conscious learning and analysis of the language being learned. Krashen argues that consciously learned language rules play a limited role in language use, serving as a monitor that could check second language output for form — assuming the learner has time, sufficient knowledge, and inclination (the monitor hypothesis). Subsequent work, by other researchers, on the interaction hypothesis and the comprehensible output hypothesis, has suggested that opportunities for output and interaction may also be necessary for learners to reach more advanced levels.

Research on how exactly learners acquire a new language spans several different areas. Focus is directed toward providing proof of whether basic linguistic skills are innate (nature), acquired (nurture), or a combination of the two attributes. Cognitive approaches to SLA research deal with the processes in the brain that underpin language acquisition, for example how paying attention to language affects the ability to learn it, or how language acquisition is related to short-term and long-term memory. Sociocultural approaches reject the notion that SLA is a purely psychological phenomenon and attempt to explain it in a social context.

Some key social factors that influence SLA are the level of immersion, connection to the L2 community, and gender. Linguistic approaches consider language separately from other kinds of knowledge and attempt to use findings from the wider

study of linguistics to explain SLA. There is also a considerable body of research about how SLA can be affected by individual factors such as age and learning strategies. A commonly discussed topic regarding age in SLA is the critical period hypothesis, which suggests that individuals lose the ability to fully learn a language after a particular age in childhood. Another topic of interest in SLA is the differences between adult and child learners. Learning strategies are commonly categorized as learning or communicative strategies and are developed to improve their respective acquisition skills. Affective factors are emotional factors that influence an individual's ability to learn a new language. Common affective factors that influence acquisition are anxiety, personality, social attitudes, and motivation.

Individuals may also lose a language through a process called second-language attrition. This is often caused by lack of use or exposure to a language over time. The severity of attrition depends on a variety of factors including level of proficiency, age, social factors, and motivation at the time of acquisition. Finally, classroom research deals with the effect that language instruction has on acquisition.

Definitions

Second language refers to any language learned in addition to a person's first language; although the concept is named*second*-language acquisition, it can also incorporate the learning of third, fourth, or subsequent languages. Second-language acquisition refers to what learners do; it does not refer to practices in language teaching, although teaching can affect acquisition. The term *acquisition* was originally used to

emphasize the non-conscious nature of the learning process, but in recent years*learning* and *acquisition* have become largely synonymous.

SLA can incorporate heritage language learning, but it does not usually incorporate bilingualism. Most SLA researchers see bilingualism as being the end result of learning a language, not the process itself, and see the term as referring to native-like fluency. Writers in fields such as education and psychology, however, often use bilingualism loosely to refer to all forms of multilingualism. SLA is also not to be contrasted with the acquisition of a foreign language; rather, the learning of second languages and the learning of foreign languages involve the same fundamental processes in different situations.

Research background

The academic discipline of second-language acquisition is a sub-discipline of applied linguistics. It is broad-based and relatively new. As well as the various branches of linguistics, second-language acquisition is also closely related to psychology and education. To separate the academic discipline from the learning process itself, the terms *second-language acquisition research*, *second-language studies*, and *second-language acquisition studies* are also used.

SLA research began as an interdisciplinary field; because of this, it is difficult to identify a precise starting date. However, two papers in particular are seen as instrumental to the development of the modern study of SLA: Pit Corder's 1967 essay *The Significance of Learners' Errors* and Larry Selinker's 1972 article *Interlanguage*. The field saw a great deal of

development in the following decades. Since the 1980s, SLA has been studied from a variety of disciplinary perspectives, and theoretical perspectives. In the early 2000s, some research suggested an equivalence between the acquisition of human languages and that of computer languages (e.g. Java) by children in the 5 to 11 year age window, though this has not been widely accepted amongst educators. Significant approaches in the field today are systemic functional linguistics, sociocultural theory, cognitive linguistics, Noam Chomsky's universal grammar, skill acquisition theory and connectionism.

There has been much debate about exactly how language is learned and many issues are still unresolved. There are many theories of second-language acquisition, but none are accepted as a complete explanation by all SLA researchers. Due to the interdisciplinary nature of the field of SLA, this is not expected to happen in the foreseeable future. Although attempts have been made to provide a more unified account that tries to bridge first language acquisition and second language learning research.

Stages

Stephen Krashen divides the process of second-language acquisition into five stages: preproduction, early production, speech emergence, intermediate fluency, and advanced fluency. The first stage, preproduction, is also known as the silent period. Learners at this stage have a receptive vocabulary of up to 500 words, but they do not yet speak their second language. Not all learners go through a silent period. Some learners start speaking straight away, although their output may consist of

imitation rather than creative language use. Others may be required to speak from the start as part of a language course. For learners that do go through a silent period, it may last around three to six months.

The second of Krashen's stages of acquisition is early production, during which learners are able to speak in short phrases of one or two words. They can also memorize chunks of language, although they may make mistakes when using them. Learners typically have both an active and receptive vocabulary of around 1000 words. This stage normally lasts for around six months.

The third stage is speech emergence. Learners' vocabularies increase to around 3000 words during this stage, and they can communicate using simple questions and phrases. They may often make grammatical errors.

The fourth stage is intermediate fluency. At this stage, learners have a vocabulary of around 6000 words, and can use more complicated sentence structures. They are also able to share their thoughts and opinions. Learners may make frequent errors with more complicated sentence structures.

The final stage is advanced fluency, which is typically reached somewhere between five and ten years of learning the language. Learners at this stage can function at a level close to native speakers.

Krashen has also developed a number of hypotheses discussing the nature of second language learners' thought processes and the development of self-awareness during second language

acquisition. The most prominent of these hypotheses are Monitor Theory and the Affective Filter hypothesis.

Language difficulty and learning time

The time taken to reach a high level of proficiency can vary depending on the language learned. In the case of native English speakers, some estimates were provided by the *Foreign Service Institute* (FSI) of the U.S. Department of State — which compiled approximate learning expectations for a number of languages for their professional staff (native English speakers who generally already know other languages). *Category I Languages* include e.g. Italian and Swedish (24 weeks or 600 class hours) and French (30 weeks or 750 class hours). *Category II Languages* include German, Haitian Creole, Indonesian, Malay, Swahili (approx. 36 weeks or 900 class hours). *Category III Languages* include a lot of languages like Finnish, Polish, Russian, Tagalog, Vietnamese and many others (approx. 44 weeks, 1100 class hours).

Of the 63 languages analyzed, the five most difficult languages to reach proficiency in speaking and reading, requiring 88 weeks (2200 class hours, *Category IV Languages*), are Arabic, Cantonese, Mandarin, Japanese, and Korean. The Foreign Service Institute and the National Virtual Translation Center both note that Japanese is typically more difficult to learn than other languages in this group.

There are other rankings of language difficulty as the one by *The British Foreign Office Diplomatic Service Language Centre*

which lists the difficult languages in Class I (Cantonese, Japanese, Korean, Mandarin); the easier languages are in Class V (e.g. Afrikaans, Bislama, Catalan, French, Spanish, Swedish).

The bottleneck hypothesis

The bottleneck hypothesis strives to identify components of grammar that are easier or more difficult to acquire than others. It argues that functional morphology is the bottleneck of language acquisition, meaning that it is more difficult than other linguistic domains such as syntax, semantics, and phonology because it combines syntactic, semantic, and phonological features that affect the meaning of a sentence. For example, knowledge of the formation of the past tense in English requires both phonological patterns such as allomorphs at the end of the verb and irregular verb forms. Article acquisition is also difficult for L1 speakers of languages without articles, such as Korean and Russian.

One study compared learner judgments of a syntactic feature, V2, and a morphological property, subject-verb agreement, using an acceptability judgment task. Researchers found that while Norwegian speakers who are intermediate and advanced learners of English could successfully assess the grammaticality of V2, they had significantly more difficulty with subject-verb agreement, which is predicted by the bottleneck hypothesis.

Cognitive and scientific reasons for the importance of this theory aside, the bottleneck hypothesis can also be of practical benefit as educators can maximize their time and focus on

difficult problems in SLA classroom settings rather than placing attention on concepts that can be grasped with relative ease.

The cumulative effects hypothesis

This hypothesis claims that second-language acquisition may impose extra difficulties on children with specific language impairment (SLI), whose language delay extends into their school years due to deficits in verbal memory and processing mechanisms in comparison to children with typical development (TD).

Existing research on individuals with SLI and bilingualism has been limited and thus there is a need for data showing how to support bilingual development in children with SLI. “Cumulative” refers to the combination of the effects of both internal deficits in language learning and external complications in input and experience caused by bilingualism, which could in turn overwhelm the learner with SLI.

The theory predicts that bilingual children with SLI will be disadvantaged, falling behind both their monolingual peers with SLI and bilingual peers with TD. Paradis' longitudinal study examined the acquisition of tense morphology over time in children with SLI who are learning English as a second language. The study found that the acquisition profile for children with SLI is similar to those reported for monolinguals with SLI and TD, showing inconsistencies with CEH. This has provided evidence that SLA will not negatively harm children with SLI and could in fact be beneficial.

Comparisons with first-language acquisition

Adults who learn a second language differ from children learning their first language in at least three ways: children are still developing their brains whereas adults have mature minds, and adults have at least a first language that orients their thinking and speaking. Although some adult second-language learners reach very high levels of proficiency, pronunciation tends to be non-native. This lack of native pronunciation in adult learners is explained by the critical period hypothesis. When a learner's speech plateaus, it is known as fossilization.

Some errors that second-language learners make in their speech originate in their first language. For example, Spanish speakers learning Englishmay say "Is raining" rather than "It is raining", leaving out the subject of the sentence. This kind of influence of the first language on the second is known as *negative*language transfer. French speakers learning English, however, do not usually make the same mistake of leaving out "it" in "It is raining." This is because pronominal and impersonal sentence subjects can be omitted (or as in this case, are not used in the first place) in Spanish but not in French. The French speaker knowing to use a pronominal sentence subject when speaking English is an example of *positive* language transfer. Not all errors occur in the same ways; even two individuals with the same native language learning the same second language still have the potential to utilize different parts of their native language. Likewise, these

same two individuals may develop near-native fluency in different forms of grammar.

Also, when people learn a second language, the way they speak their first language changes in subtle ways. These changes can be with any aspect of language, from pronunciation and syntax to the gestures the learner makes and the language features they tend to notice. For example, French speakers who spoke English as a second language pronounced the /t/ sound in French differently from monolingual French speakers. This kind of change in pronunciation has been found even at the onset of second-language acquisition; for example, English speakers pronounced the English /p t k/ sounds, as well as English vowels, differently after they began to learn Korean. These effects of the second language on the first led Vivian Cook to propose the idea of multi-competence, which sees the different languages a person speaks not as separate systems, but as related systems in their mind.

Learner language

Learner language is the written or spoken language produced by a learner. It is also the main type of data used in second-language acquisition research. Much research in second-language acquisition is concerned with the internal representation of a language in the mind of the learner, and in how those representations change over time. It is not yet possible to inspect these representations directly with brain scans or similar techniques, so SLA researchers are forced to make inferences about these rules from learners' speech or writing.

Interlanguage

Originally, attempts to describe learner language were based on comparing different languages and on analyzing learners' errors. However, these approaches were unable to predict all the errors that learners made when in the process of learning a second language. For example, Serbo-Croat speakers learning English may say "What does Pat doing now?", although this is not a valid sentence in either language.

Additionally, Yip found that ergative verbs in English are regularly mis-passivized by L2 learners of English whose first language is Mandarin. For instance, even advanced learners may form utterances such as "what was happened?" despite the fact that this construction has no obvious source in neither L1 nor L2. This could be because L2 speakers interpret ergatives as transitive, as these are the only types of verbs that allow passivization in English.

To explain this kind of systematic error, the idea of the *interlanguage*was developed. An interlanguage is an emerging language system in the mind of a second-language learner. A learner's interlanguage is not a deficient version of the language being learned filled with random errors, nor is it a language purely based on errors introduced from the learner's first language. Rather, it is a language in its own right, with its own systematic rules. It is possible to view most aspects of language from an interlanguage perspective, including grammar, phonology, lexicon, and pragmatics.

There are three different processes that influence the creation of interlanguages:

- *Language transfer*. Learners fall back on their mother tongue to help create their language system. Transfer can be positive, i.e. promote learning, or negative, i.e. lead to mistakes. In the latter case, linguists also use the term interference error.
- *Overgeneralization*. Learners use rules from the second language in roughly the same way that children overgeneralise in their first language. For example, a learner may say "I goed home", overgeneralizing the English rule of adding *-ed* to create past tense verb forms. English children also produce forms like *goed*, *sticked*, and *bringed*. German children equally overextend regular past tense forms to irregular forms.
- *Simplification*. Learners use a highly simplified form of language, similar to speech by children or in pidgins. This may be related to linguistic universals.

The concept of interlanguage has become very widespread in SLA research, and is often a basic assumption made by researchers.

Sequences in the acquisition of English inflectional morphology

In the 1970s, several studies investigated the order in which learners acquired different grammatical structures. These studies showed that there was little change in this order among learners with different first languages. Furthermore, it

showed that the order was the same for adults and children, and that it did not even change if the learner had language lessons. This supported the idea that there were factors other than language transfer involved in learning second languages, and was a strong confirmation of the concept of interlanguage.

However, the studies did not find that the orders were exactly the same. Although there were remarkable similarities in the order in which all learners learned second-language grammar, there were still some differences among individuals and among learners with different first languages. It is also difficult to tell when exactly a grammatical structure has been learned, as learners may use structures correctly in some situations but not in others. Thus it is more accurate to speak of *sequences* of acquisition, in which specific grammatical features in a language are acquired before or after certain others but the overall order of acquisition is less rigid. For example, if neither feature B nor feature D can be acquired until feature A has been acquired (feature B and D depend on A) and feature C depends on B, but D does not depend on B (or, therefore, on C), then acquisition orders (A, B, C, D) and (A, D, B, C) are possible, as they are both valid topological orderings.

Learnability and teachability

Learnability has emerged as a theory explaining developmental sequences that crucially depend on learning principles, which are viewed as fundamental mechanisms of language acquisition within learnability theory. Some examples of learning principles include the uniqueness principle and the subset principle. The uniqueness principle refers to learners' preference for one-to-one mapping between form and meaning,

while the subset principle posits that learners are conservative in that they begin with the narrowest hypothesis space that is compatible with available data. Both of these principles have been used to explain children's ability to evaluate grammaticality in spite of the lack of explicit negative evidence. They have also been used to explain errors in SLA, as the creation of supersets could signal over-generalization, causing acceptance or production of ungrammatical sentences.

Pienemann'steachability hypothesis is based on the idea that there is a hierarchy on stages of acquisition and instruction in SLA should be compatible to learners' current acquisitional status. Recognizing learners' developmental stages is important as it enables teachers to predict and classify learning errors. This hypothesis predicts that L2 acquisition can only be promoted when learners are ready to acquire given items in a natural context. One goal of learnability theory is to figure out which linguistic phenomena are susceptible to fossilization, wherein some L2 learners continue to make errors in spite of the presence of relevant input.

Variability

Although second-language acquisition proceeds in discrete sequences, it does not progress from one step of a sequence to the next in an orderly fashion. There can be considerable variability in features of learners' interlanguage while progressing from one stage to the next. For example, in one study by Rod Ellis, a learner used both "No look my card" and "Don't look my card" while playing a game of bingo. A small fraction of variation in interlanguage is *free variation*, when the learner uses two forms interchangeably. However, most

variation is *systemic variation*, variation that depends on the context of utterances the learner makes. Forms can vary depending on linguistic context, such as whether the subject of a sentence is a pronoun or a noun; they can vary depending on social context, such as using formal expressions with superiors and informal expressions with friends; and also, they can vary depending on psycholinguistic context, or in other words, on whether learners have the chance to plan what they are going to say. The causes of variability are a matter of great debate among SLA researchers.

Language transfer

One important difference between first-language acquisition and second-language acquisition is that the process of second-language acquisition is influenced by languages that the learner already knows. This influence is known as *language transfer*. Language transfer is a complex phenomenon resulting from interaction between learners' prior linguistic knowledge, the target-language input they encounter, and their cognitive processes. Language transfer is not always from the learner's native language; it can also be from a second language, or a third. Neither is it limited to any particular domain of language; language transfer can occur in grammar, pronunciation, vocabulary, discourse, and reading.

Language transfer often occurs when learners sense a similarity between a feature of a language they already know and a feature of the interlanguage they have developed. If this happens, the acquisition of more complicated language forms may be delayed in favor of simpler language forms that resemble those of the language the learner is familiar with.

Learners may also decline to use some language forms at all if they are perceived as being too distant from their first language.

Language transfer has been the subject of several studies, and many aspects of it remain unexplained. Various hypotheses have been proposed to explain language transfer, but there is no single widely accepted explanation of why it occurs.

Some linguists prefer to use cross-linguistic influence to describe this phenomenon. Studies on bilingual children find bidirectional cross-linguistic influence; for example, Nicoladis (2012) reported that bilingual children aged three to four produce French-like periphrastic constructions e.g. "the hat of the dog" and ungrammatical English-like reversed possessive structures e.g. "*chien chapeau*" (dog hat) significantly more than their monolingual peers.

Though periphrastic constructions are expected as they are grammatical in both English and French, reversed possessives in French are ungrammatical and thus unexpected.

In a study exploring cross-linguistic influence in word order by comparing Dutch-English bilingual and English monolingual children, Unsworth found that bilingual children were more likely to accept incorrect V2 word orders in English than monolinguals with both auxiliary and main verbs. Dominance was a predictor of this phenomenon; Dutch-dominant children showed less sensitivity to word order than English-dominant ones, though this effect was small and there was individual variation.

Language dominance

The term language dominance can be defined in terms of differences in frequency of use and differences in proficiency in bilinguals. How basic or advanced a speaker's L2 level will be is determined by a complex range of environmental, individual and other factors.

Language dominance may change over time through the process of language attrition, in which some L2 skills begin to match or even overtake those of L1. Research suggests a correlation between amount of language exposure and cross-linguistic influence; language dominance is considered to have an impact on the direction of transfer.

One study found that transfer is asymmetrical and predicted by dominance, as Cantonese dominant children showed clear syntactic transfer in many areas of grammar from Cantonese to English but not vice versa. MLU, mean length of utterance, is a common measurement of linguistic productivity and language dominance in children.

Input and interaction

The primary factor affecting language acquisition appears to be the input that the learner receives. Stephen Krashen took a very strong position on the importance of input, asserting that comprehensible input is all that is necessary for second-language acquisition. Krashen pointed to studies showing that the length of time a person stays in a foreign country is closely linked with their level of language acquisition. Further evidence for input comes from studies on reading: large

amounts of free voluntary reading have a significant positive effect on learners' vocabulary, grammar, and writing. Input is also the mechanism by which people learn languages according to the universal grammar model.

The type of input may also be important. One tenet of Krashen's theory is that input should not be grammatically sequenced. He claims that such sequencing, as found in language classrooms where lessons involve practicing a "structure of the day", is not necessary, and may even be harmful.

While input is of vital importance, Krashen's assertion that *only* input matters in second-language acquisition has been contradicted by more recent research. For example, students enrolled in French-language immersion programs in Canada still produced non-native-like grammar when they spoke, even though they had years of meaning-focused lessons and their listening skills were statistically native-level. Output appears to play an important role, and among other things, can help provide learners with feedback, make them concentrate on the form of what they are saying, and help them to automatize their language knowledge. These processes have been codified in the theory of comprehensible output.

Researchers have also pointed to interaction in the second language as being important for acquisition.

According to Long's interaction hypothesis the conditions for acquisition are especially good when interacting in the second language; specifically, conditions are good when a breakdown in communication occurs and learners must negotiate for meaning. The modifications to speech arising from interactions

like this help make input more comprehensible, provide feedback to the learner, and push learners to modify their speech.

Factors and approaches to SLA

Cognitive factors

Much modern research in second-language acquisition has taken a cognitive approach. Cognitive research is concerned with the mental processes involved in language acquisition, and how they can explain the nature of learners' language knowledge.

This area of research is based in the more general area of cognitive science, and uses many concepts and models used in more general cognitive theories of learning. As such, cognitive theories view second-language acquisition as a special case of more general learning mechanisms in the brain. This puts them in direct contrast with linguistic theories, which posit that language acquisition uses a unique process different from other types of learning. The dominant model in cognitive approaches to second-language acquisition, and indeed in all second-language acquisition research, is the computational model. The computational model involves three stages. In the first stage, learners retain certain features of the language input in short-term memory. (This retained input is known as *intake*.) Then, learners convert some of this intake into second-language knowledge, which is stored in long-term memory. Finally, learners use this second-language knowledge to produce spoken output. Cognitive theories attempt to codify

both the nature of the mental representations of intake and language knowledge, and the mental processes that underlie these stages.

In the early days of second-language acquisition research on interlanguage was seen as the basic representation of second-language knowledge; however, more recent research has taken a number of different approaches in characterizing the mental representation of language knowledge. There are theories that hypothesize that learner language is inherently variable, and there is the functionalist perspective that sees acquisition of language as intimately tied to the function it provides. Some researchers make the distinction between *implicit* and *explicit* language knowledge, and some between *declarative* and *procedural* language knowledge. There have also been approaches that argue for a *dual-mode system* in which some language knowledge is stored as rules, and other language knowledge as items.

The mental processes that underlie second-language acquisition can be broken down into micro-processes and macro-processes. Micro-processes include attention; working memory; integration and restructuring. Restructuring is the process by which learners change their interlanguage systems; and *monitoring* is the conscious attending of learners to their own language output. Macro-processes include the distinction between intentional learning and incidental learning; and also the distinction between explicit and implicit learning. Some of the notable cognitive theories of second-language acquisition include the nativization model, the multidimensional model and processability theory, emergentist models, the competition model, and skill-acquisition theories.

Other cognitive approaches have looked at learners' speech production, particularly learners' speech planning and communication strategies. Speech planning can have an effect on learners' spoken output, and research in this area has focused on how planning affects three aspects of speech: complexity, accuracy, and fluency.

Of these three, planning effects on fluency has had the most research attention. Communication strategies are conscious strategies that learners employ to get around any instances of communication breakdown they may experience. Their effect on second-language acquisition is unclear, with some researchers claiming they help it, and others claiming the opposite.

An important idea in recent cognitive approaches is the way that learning itself changes over development. For example, connectionist models that explain L1 language phenomena in different languages (e.g., Japanese, English) can also be used to develop L2 models by first training on the L1 (e.g., Korean) and then training on the L2 (e.g. English). By using different learning rates for syntax and lexical learning that change over development, the model can explain sensitive period effects and differences in the effect of language exposure on different types of learners.

Sociocultural factors

From the early days of the discipline researchers have also acknowledged that social aspects play an important role. There have been many different approaches to sociolinguistic study of second-language acquisition, and indeed, according to Rod Ellis, this plurality has meant that "sociolinguistic SLA is

replete with a bewildering set of terms referring to the social aspects of L2 acquisition". Common to each of these approaches, however, is a rejection of language as a purely psychological phenomenon; instead, sociolinguistic research views the social context in which language is learned as essential for a proper understanding of the acquisition process.

Ellis identifies three types of social structure that affect acquisition of second languages: sociolinguistic setting, specific social factors, and situational factors. Sociolinguistic setting refers to the role of the second language in society, such as whether it is spoken by a majority or a minority of the population, whether its use is widespread or restricted to a few functional roles, or whether the society is predominantly bilingual or monolingual. Ellis also includes the distinction of whether the second language is learned in a natural or an educational setting. Specific social factors that can affect second-language acquisition include age, gender, social class, and ethnic identity, with ethnic identity being the one that has received most research attention. Situational factors are those that vary between each social interaction. For example, a learner may use more polite language when talking to someone of higher social status, but more informal language when talking with friends.

Immersion programs provide a sociolinguistic setting that facilitates second-language acquisition. Immersion programs are educational programs where children are instructed in an L2 language. Although the language of instruction is the L2 language, the curriculum parallels that of non-immersion programs and clear support exists in the L1 language, as the teachers are all bilingual.

The goal of these programs is to develop a high level of proficiency in both the L1 and L2 languages. Students in immersion programs have been shown to have greater levels of proficiency in their second language than students who receive second language education only as a subject in school. This is especially true in terms of their receptive skills. Also, students who join immersion programs earlier generally have greater second-language proficiency than their peers who join later. However, students who join later have been shown to gain native-like proficiency. Although immersion students' receptive skills are especially strong, their productive skills may suffer if they spend the majority of their time listening to instruction only. Grammatical skills and the ability to have precise vocabulary are particular areas of struggle. It is argued that immersion is necessary, but not sufficient for the development of native-like proficiency in a second language. Opportunities to engage in sustained conversation, and assignments that encourage syntactical, as well as semantic development help develop the productive skills necessary for bilingual proficiency.

A learner's sense of connection to their in-group, as well as to the community of the target language emphasize the influence of the sociolinguistic setting, as well as social factors within the second-language acquisition process. Social Identity Theory argues that an important factor for second language acquisition is the learner's perceived identity in relation to the community of the language being learned, as well as how the community of the target language perceives the learner. Whether or not a learner feels a sense of connection to the community or culture of the target language helps determine their social distance from the target culture. A smaller social

distance is likely to encourage learners to acquire the second language, as their investment in the learning process is greater. Conversely, a greater social distance discourages attempts to acquire the target language. However, negative views not only come from the learner, but the community of the target language might feel greater social distance to the learner, limiting the learner's ability to learn the language. Whether or not bilingualism is valued by the culture or community of the learner is an important indicator for the motivation to learn a language.

Gender, as a social factor, also influences SLA. Females have been found to have higher motivation and more positive attitudes than males for second-language acquisition. However, females are also more likely to present higher levels of anxiety, which may inhibit their ability to efficiently learn a new language.

There have been several models developed to explain social effects on language acquisition. Schumann's Acculturation Model proposes that learners' rate of development and ultimate level of language achievement is a function of the "social distance" and the "psychological distance" between learners and the second-language community. In Schumann's model the social factors are most important, but the degree to which learners are comfortable with learning the second language also plays a role. Another sociolinguistic model is Gardner's socio-educational model, which was designed to explain classroom language acquisition. Gardner's model focuses on the emotional aspects of SLA, arguing that positive motivation contributes to an individuals willingness to learn L2; furthermore, the goal of an individual to learn a L2 is based on

the idea that the individual has a desire to be part of a culture, in other words, part of a (the targeted language) mono-linguistic community. Factors, such as *integrativeness* and *attitudes towards the learning situation* drive motivation. The outcome of positive motivation is not only linguistic, but non-linguistic, such that the learner has met the desired goal. Although there are many critics of Gardner's model, nonetheless many of these critics have been influenced by the merits that his model holds. The inter-group model proposes "ethnolinguistic vitality" as a key construct for second-language acquisition. Language socialization is an approach with the premise that "linguistic and cultural knowledge are *constructed* through each other", and saw increased attention after the year 2000. Finally, Norton's theory of social identity is an attempt to codify the relationship between power, identity, and language acquisition.

A unique approach to SLA is Sociocultural theory. It was originally developed by Lev Vygotsky and his followers. Central to Vygotsky's theory is the concept of a zone of proximal development (ZPD). The ZPD notion states that social interaction with more advanced target language users allows one to learn language at a higher level than if they were to learn language independently. Sociocultural theory has a fundamentally different set of assumptions to approaches to second-language acquisition based on the computational model. Furthermore, although it is closely affiliated with other social approaches, it is a theory of mind and not of general social explanations of language acquisition. According to Ellis, "It is important to recognize... that this paradigm, despite the label 'sociocultural' does not seek to explain how learners acquire the cultural values of the L2 but rather how knowledge

of an L2 is internalized through experiences of a sociocultural nature."

Linguistic factors

Linguistic approaches to explaining second-language acquisition spring from the wider study of linguistics. They differ from cognitive approaches and sociocultural approaches in that they consider linguistic knowledge to be unique and distinct from any other type of knowledge. The linguistic research tradition in second-language acquisition has developed in relative isolation from the cognitive and sociocultural research traditions, and as of 2010 the influence from the wider field of linguistics was still strong. Two main strands of research can be identified in the linguistic tradition: generative approaches informed by universal grammar, and typological approaches.

Typological universals are principles that hold for all the world's languages. They are found empirically, by surveying different languages and deducing which aspects of them could be universal; these aspects are then checked against other languages to verify the findings. The interlanguages of second-language learners have been shown to obey typological universals, and some researchers have suggested that typological universals may constrain interlanguage development.

The theory of universal grammar was proposed by Noam Chomsky in the 1950s, and has enjoyed considerable popularity in the field of linguistics. It focuses on describing the linguistic competence of an individual. He believed that

children not only acquire language by learning descriptive rules of grammar; he claimed that children *creatively* play and form words as they learn language, creating meaning of these words, as opposed to the mechanism of memorizing language. It consists of a set of *principles*, which are universal and constant, and a set of *parameters*, which can be set differently for different languages. The "universals" in universal grammar differ from typological universals in that they are a mental construct derived by researchers, whereas typological universals are readily verifiable by data from world languages. It is widely accepted among researchers in the universal grammar framework that all first-language learners have access to universal grammar; this is not the case for second-language learners, however, and much research in the context of second-language acquisition has focused on what level of access learners may have. there is ongoing debate among generative linguists surrounding whether L2-users have full or partial access to universal grammar. This can be seen through acceptability judgment tests. For example, one study found that during a comprehension task, while English L1 speakers learning Spanish may accept the imperfect aspect in appropriate conditions, even at higher levels of proficiency, they do not reject the use of the Preterite tense in continuous and habitual contexts.

Universal grammar theory can account for some of the observations of SLA research. For example, L2-users often display knowledge about their L2 that they have not been exposed to. L2-users are often aware of ambiguous or ungrammatical L2 units that they have not learned from any external source, nor from their pre-existing L1 knowledge. This unsourced knowledge suggests the existence of a universal

grammar. Another piece of evidence that generative linguists tend to use is the poverty of the stimulus, which states that children acquiring language lack sufficient data to fully acquire all facets of grammar in their language, causing a mismatch between input and output.

The fact that children are only exposed to positive evidence yet have intuition about which word strings are ungrammatical may also be indicative of universal grammar. However, L2 learners have access to negative evidence as they are explicitly taught about ungrammaticality through corrections or grammar teaching.

Individual variation

There is considerable variation in the rate at which people learn second languages, and in the language level that they ultimately reach. Some learners learn quickly and reach a near-native level of competence, but others learn slowly and get stuck at relatively early stages of acquisition, despite living in the country where the language is spoken for several years. The reason for this disparity was first addressed with the study of language learning aptitude in the 1950s, and later with the *good language learner studies* in the 1970s.

More recently research has focused on a number of different factors that affect individuals' language learning, in particular strategy use, social and societal influences, personality, motivation, and anxiety. The relationship between age and the ability to learn languages has also been a subject of long-standing debate.

Age

The issue of age was first addressed with the critical period hypothesis. The strict version of this hypothesis states that there is a cut-off age at about 12, after which learners lose the ability to fully learn a language. However, the exact age marking the end of the critical period is debated, and ranges from age 6 to 13, with many arguing that it is around the onset of puberty. This strict version has since been rejected for second-language acquisition, as some adult and adolescent learners have been observed who reach native-like levels of pronunciation and general fluency faster than young children. However, in general, adolescent and adult learners of a second-language rarely achieve the native-like fluency that children who acquire both languages from birth display, despite often progressing faster in the initial stages. This has led to speculation that age is indirectly related to other, more central factors that affect language learning.

Children who acquire two languages from birth are called simultaneous bilinguals. In these cases, both languages are spoken to the children by their parents or caregivers and they grow up knowing the two languages. These children generally reach linguistic milestones at the same time as their monolingual peers. Children who do not learn two languages from infancy, but learn one language from birth, and another at some point during childhood, are referred to as sequential bilinguals. People often assume that a sequential bilingual's first language is their most proficient language, but this is not always the case. Over time and experience, a child's second language may become his or her strongest. This is especially likely to happen if a child's first language is a minority

language spoken at home, and the child's second language is the majority language learned at school or in the community before the age of five. Proficiency for both simultaneous and sequential bilinguals is dependent upon the child's opportunities to engage in meaningful conversations in a variety of contexts.

Often simultaneous bilinguals are more proficient in their languages than sequential bilinguals. One argument for this is that simultaneous bilinguals develop more distinct representations of their languages, especially with regards tophonological and semantic levels of processing. This would cause learners to have more differentiation between the languages, leading them to be able to recognize the subtle differences between the languages that less proficient learners would struggle to recognize. Learning a language earlier in life would help develop these distinct representations of language, as the learner's first language would be less established. Conversely, learning a language later in life would lead to more similar semantic representations.

Although child learners more often acquire native-like proficiency, older child and adult learners often progress faster in the initial stages of learning. Older child and adult learners are quicker at acquiring the initial grammar knowledge than child learners, however, with enough time and exposure to the language, children surpass their older peers. Once surpassed, older learners often display clear language deficiencies compared to child learners. This has been attributed to having a solid grasp on the first language or mother tongue they were first immersed into. Having this cognitive ability already developed can aid the process of learning a second language

since there is a better understanding of how language works. For this same reason interaction with family and further development of the first language is encouraged along with positive reinforcement. The exact language deficiencies that occur past a certain age are not unanimously agreed upon. Some believe that only pronunciation is affected, while others believe other abilities are affected as well. However, some differences that are generally agreed upon include older learners having a noticeable accent, a smaller vocabulary, and making several linguistic errors.

One explanation for this difference in proficiency between older learners and younger learners involves Universal Grammar. Universal Grammar is a debated theory that suggests that people have innate knowledge of universal linguistic principles that is present from birth. These principles guide children as they learn a language, but its parameters vary from language to language. The theory assumes that, while Universal Grammar remains into adulthood, the ability to reset the parameters set for each language is lost, making it more difficult to learn a new language proficiently. Since older learners would already have an established native language, the language acquisition process is very different for them, than young learners. The rules and principles that guide the use of the learners' native language plays a role in the way the second language is developed.

Some nonbiological explanations for second-language acquisition age differences include variations in social and psychological factors, such as motivation; the learner's linguistic environment; and the level of exposure. Even with less advantageous nonbiological influences, many young

children attain a greater level of proficiency in their second language than older learners with more advantageous nonbiologicalinfluences.

Strategies

Considerable attention has been paid to the strategies learners use to learn a second language. Strategies have been found to be of critical importance, so much so that *strategic competence* has been suggested as a major component of communicative competence. Strategies are commonly divided into *learning strategies* and *communicative strategies*, although there are other ways of categorizing them. Learning strategies are techniques used to improve learning, such as mnemonics or using a dictionary. Communicative strategies are strategies a learner uses to convey meaning even when he or she doesn't have access to the correct form, such as using pro-forms like *thing*, or using non-verbal means such as gestures. If learning strategies and communicative strategies are used properly language acquisition is successful. Some points to keep in mind while learning an additional language are: providing information that is of interest to the student, offering opportunities for the student to share their knowledge and teaching appropriate techniques for the uses of the learning resources available.

Another strategy may include intentional ways to acquire or improve their second language skills. Adult immigrants and/or second language learners seeking to acquire a second language can engage in different activities to receive and share knowledge as well as improve their learning; some of these include:

- incidental or informal learning (media resources, family/friend interactions, work interactions)
- purposeful learning (self-study, taking language classes)
- pursuing formal education

Affective factors

The learner's attitude to the learning process has also been identified as being critically important to second-language acquisition. Anxiety in language-learning situations has been almost unanimously shown to be detrimental to successful learning. Anxiety interferes with the mental processing of language because the demands of anxiety-related thoughts create competition for mental resources. This results in less available storage and energy for tasks required for language processing. Not only this, but anxiety is also usually accompanied by self-deprecating thoughts and fear of failure, which can be detrimental to an individual's ability to learn a new language. Learning a new language provides a unique situation that may even produce a specific type of anxiety, called language anxiety, that affects the quality of acquisition. Also, anxiety may be detrimental for SLA because it can influence a learner's ability to attend to, concentrate on, and encode language information. It may affect speed and accuracy of learning. Further, the apprehension created as a result of anxiety inhibits the learner's ability to retrieve and produce the correct information.

A related factor, personality, has also received attention. There has been discussion about the effects of extravert and introvert

personalities. Extraverted qualities may help learners seek out opportunities and people to assist with L2 learning, whereas introverts may find it more difficult to seek out such opportunities for interaction. However, it has also been suggested that, while extraverts might experience greater fluency, introverts are likely to make fewer linguistic errors. Further, while extraversion might be beneficial through its encouragement of learning autonomously, it may also present challenges as learners may find reflective and time-management skills to be difficult. However, one study has found that there were no significant differences between extraverts and introverts on the way they achieve success in a second language.

Other personality factors, such as conscientiousness, agreeableness, and openness influence self-regulation, which helps L2 learners engage, process meaning, and adapt their thoughts, feelings, and actions to benefit the acquisition process. SLA research has shown conscientiousness to be associated with time-management skills, metacognition, analytic learning, and persistence; agreeableness to effort; and openness to elaborative learning, intelligence, and metacognition. Both genetics and the learner's environment impact the personality of the learner, either facilitating or hindering an individual's ability to learn.

Social attitudes such as gender roles and community views toward language learning have also proven critical. Language learning can be severely hampered by cultural attitudes, with a frequently cited example being the difficulty of Navajo children in learning English.

Also, the motivation of the individual learner is of vital importance to the success of language learning. Motivation is influenced by goal salience, valence, and self-efficacy. In this context, goal salience is the importance of the L2 learner's goal, as well as how often the goal is pursued; valence is the value the L2 learner places on SLA, determined by desire to learn and attitudes about learning the L2; and self-efficacy is the learner's own belief that he or she is capable of achieving the linguistic goal. Studies have consistently shown that *intrinsic motivation*, or a genuine interest in the language itself, is more effective over the long term than *extrinsic motivation*, as in learning a language for a reward such as high grades or praise. However, motivation is dynamic and, as a L2 learner's fluency develops, their extrinsic motivation may evolve to become more intrinsic. Learner motivation can develop through contact with the L2 community and culture, as learners often desire to communicate and identify with individuals in the L2 community. Further, a supportive learning environment facilitates motivation through the increase in self-confidence and autonomy. Learners in a supportive environment are more often willing to take on challenging tasks, thus encouraging L2 development.

Attrition

Attrition is the loss of proficiency in a language caused by a lack of exposure to or use of a language. It is a natural part of the language experience as it exists within a dynamic environment. As the environment changes, the language adapts. One way it does this is by using L1 as a tool to navigate the periods of change associated with acquisition and

attrition. A learner's L2 is not suddenly lost with disuse, but its communicative functions are slowly replaced by those of the L1.

Similar to second-language acquisition, second-language attrition occurs in stages. However, according to the regression hypothesis, the stages of attrition occur in reverse order of acquisition. With acquisition, receptive skills develop first, and then productive skills, and with attrition, productive skills are lost first, and then receptive skills.

Age, proficiency level, and social factors play a role in the way attrition occurs. Most often younger children are quicker than adults to lose their L2 when it is left unused. However, if a child has established a high level of proficiency, it may take them several years to lose the language. Proficiency level seems to play the largest role in the extent of attrition. For very proficient individuals, there is a period of time where very little, if any, attrition is observed. For some, residual learning might even occur, which is the apparent improvement within the L2. Within the first five years of language disuse, the total percentage of language knowledge lost is less for a proficient individual than for someone less proficient. A cognitive psychological explanation for this suggests that a higher level of proficiency involves the use of schemas, or mental representations for linguistic structures. Schemas involve deeper mental processes for mental retrieval that are resistant to attrition. As a result, information that is tied to this system is less likely to experience less extreme attrition than information that is not. Finally, social factors may play an indirect role in attrition. In particular, motivation and attitude influence the process. Higher levels of motivation, and a

positive attitude toward the language and the corresponding community may lessen attrition. This is likely due to the higher level of competence achieved in L2 when the learner is motivated and has a positive attitude.

Classroom second-language acquisition

While considerable SLA research has been devoted to language learning in a natural setting, there have also been efforts made to investigate second-language acquisition in the classroom. This kind of research has a significant overlap with language education, and it is mainly concerned with the effect that instruction has on the learner. It also explores what teachers do, the classroom context, the dynamics of classroom communication. It is both qualitative and quantitative research.

The research has been wide-ranging. There have been attempts made to systematically measure the effectiveness of language teaching practices for every level of language, from phonetics to pragmatics, and for almost every current teaching methodology. This research has indicated that many traditional language-teaching techniques are extremely inefficient. Cited in Ellis 1994It is generally agreed that pedagogy restricted to teaching grammar rules and vocabulary lists does not give students the ability to use the L2 with accuracy and fluency. Rather, to become proficient in the second language, the learner must be given opportunities to use it for communicative purposes.

Another area of research has been on the effects of corrective feedback in assisting learners. This has been shown to vary depending on the technique used to make the correction, and the overall focus of the classroom, whether on formal accuracy or on communication of meaningful content.

There is also considerable interest in supplementing published research with approaches that engage language teachers in action research on learner language in their own classrooms. As teachers become aware of the features of learner language produced by their students, they can refine their pedagogical intervention to maximize interlanguage development.

If one wishes to acquire a language in a classroom setting only, one needs to consider the category language one wishes to acquire; the category of the desired language will determine how many hours or weeks to devote to study.

There are three main categories of languages. Category I languages are “cognate languages” like French, Spanish, and Swedish; category II languages are Finnish, Russian, and Vietnamese; category III languages are Arabic, Chinese, Japanese, and Korean. As such, the languages are categorized by their similarity to English. Respectively, category I languages require 24 weeks or 600 classroom hours to achieve proficiency; category II languages require 44 weeks or 1,100 hours; category III languages require 88 weeks or 2,200 hours .

Moreover, one can achieve proficiency in a foreign language in a classroom setting so long as one acknowledges the time commitment necessary.

Chapter 2

Phonetics and Phonology are the Study of Speech Sounds

Phonetics

Phonetics is a branch of linguistics that studies how humans produce and perceive sounds, or in the case of sign languages, the equivalent aspects of sign. Phoneticians—linguists who specialize in phonetics—study the physical properties of speech. The field of phonetics is traditionally divided into three sub-disciplines based on the research questions involved such as how humans plan and execute movements to produce speech (articulatory phonetics), how different movements affect the properties of the resulting sound (acoustic phonetics), or how humans convert sound waves to linguistic information (auditory phonetics). Traditionally, the minimal linguistic unit of phonetics is the phone—a speech sound in a language—which differs from the phonological unit of phoneme; the phoneme is an abstract categorization of phones.

Phonetics broadly deals with two aspects of human speech: production—the ways humans make sounds—and perception—the way speech is understood. The communicative modality of a language describes the method by which a language produces and perceives languages. Languages with oral-aural modalities such as English produce speech orally (using the mouth) and perceive speech aurally (using the ears). Sign languages, such as Auslan and ASL, have a manual-visual modality, producing

speech manually (using the hands) and perceiving speech visually (using the eyes). ASL and some other sign languages have in addition a manual-manual dialect for use in tactile signing by deafblind speakers where signs are produced with the hands and perceived with the hands as well.

Language production consists of several interdependent processes which transform a non-linguistic message into a spoken or signed linguistic signal. After identifying a message to be linguistically encoded, a speaker must select the individual words—known as lexical items—to represent that message in a process called lexical selection. During phonological encoding, the mental representation of the words are assigned their phonological content as a sequence of phonemes to be produced. The phonemes are specified for articulatory features which denote particular goals such as closed lips or the tongue in a particular location. These phonemes are then coordinated into a sequence of muscle commands that can be sent to the muscles, and when these commands are executed properly the intended sounds are produced.

These movements disrupt and modify an airstream which results in a sound wave. The modification is done by the articulators, with different places and manners of articulation producing different acoustic results. For example, the words *tack* and *sack* both begin with alveolar sounds in English, but differ in how far the tongue is from the alveolar ridge. This difference has large effects on the air stream and thus the sound that is produced. Similarly, the direction and source of the airstream can affect the sound. The most common

airstream mechanism is pulmonic—using the lungs—but the glottis and tongue can also be used to produce airstreams.

Language perception is the process by which a linguistic signal is decoded and understood by a listener. In order to perceive speech the continuous acoustic signal must be converted into discrete linguistic units such as phonemes, morphemes, and words. In order to correctly identify and categorize sounds, listeners prioritize certain aspects of the signal that can reliably distinguish between linguistic categories. While certain cues are prioritized over others, many aspects of the signal can contribute to perception. For example, though oral languages prioritize acoustic information, the McGurk effect shows that visual information is used to distinguish ambiguous information when the acoustic cues are unreliable.

Modern phonetics has three main branches:

- Articulatory phonetics which studies the way sounds are made with the articulators
- Acoustic phonetics which studies the acoustic results of different articulations
- Auditory phonetics which studies the way listeners perceive and understand linguistic signals.

History

Antiquity

The first known phonetic studies were carried out as early as the 6th century BCE by Sanskrit grammarians. The Hindu

scholar Pāṇiniis among the most well known of these early investigators, whose four-part grammar, written around 350 BCE, is influential in modern linguistics and still represents "the most complete generative grammar of any language yet written". His grammar formed the basis of modern linguistics and described several important phonetic principles, including voicing. This early account described resonance as being produced either by tone, when vocal folds are closed, or noise, when vocal folds are open. The phonetic principles in the grammar are considered "primitives" in that they are the basis for his theoretical analysis rather than the objects of theoretical analysis themselves, and the principles can be inferred from his system of phonology.

Modern

Advancements in phonetics after Pāṇini and his contemporaries were limited until the modern era, save some limited investigations by Greek and Roman grammarians. In the millennia between Indic grammarians and modern phonetics, the focus shifted from the difference between spoken and written language, which was the driving force behind Pāṇini's account, and began to focus on the physical properties of speech alone. Sustained interest in phonetics began again around 1800 CE with the term "phonetics" being first used in the present sense in 1841. With new developments in medicine and the development of audio and visual recording devices, phonetic insights were able to use and review new and more detailed data. This early period of modern phonetics included the development of an influential phonetic alphabet based on articulatory positions by Alexander Melville Bell. Known as

visible speech, it gained prominence as a tool in the oral education of deaf children.

Before the widespread availability of audio recording equipment, phoneticians relied heavily on a tradition of practical phonetics to ensure that transcriptions and findings were able to be consistent across phoneticians. This training involved both ear training—the recognition of speech sounds—as well as production training—the ability to produce sounds. Phoneticians were expected to learn to recognize by ear the various sounds on the International Phonetic Alphabet and the IPA still tests and certifies speakers on their ability to accurately produce the phonetic patterns of English (though they have discontinued this practice for other languages). As a revision of his visible speech method, Melville Bell developed a description of vowels by height and backness resulting in 9cardinal vowels. As part of their training in practical phonetics, phoneticians were expected to learn to produce these cardinal vowels in order to anchor their perception and transcription of these phones during fieldwork. This approach was critiqued by Peter Ladefoged in the 1960s based on experimental evidence where he found that cardinal vowels were auditory rather than articulatory targets, challenging the claim that they represented articulatory anchors by which phoneticians could judge other articulations.

Production

Language production consists of several interdependent processes which transform a nonlinguistic message into a spoken or signed linguistic signal. Linguists debate whether the process of language production occurs in a series of stages

(serial processing) or whether production processes occur in parallel. After identifying a message to be linguistically encoded, a speaker must select the individual words—known as lexical items—to represent that message in a process called lexical selection. The words are selected based on their meaning, which in linguistics is called semantic information. Lexical selection activates the word's lemma, which contains both semantic and grammatical information about the word.

After an utterance has been planned, it then goes through phonological encoding. In this stage of language production, the mental representation of the words are assigned their phonological content as a sequence of phonemes to be produced. The phonemes are specified for articulatory features which denote particular goals such as closed lips or the tongue in a particular location. These phonemes are then coordinated into a sequence of muscle commands that can be sent to the muscles, and when these commands are executed properly the intended sounds are produced. Thus the process of production from message to sound can be summarized as the following sequence:

- Message planning
- Lemma selection
- Retrieval and assignment of phonological word forms
- Articulatory specification
- Muscle commands
- Articulation
- Speech sounds

Place of articulation

Sounds which are made by a full or partial constriction of the vocal tract are called consonants. Consonants are pronounced in the vocal tract, usually in the mouth, and the location of this constriction affects the resulting sound. Because of the close connection between the position of the tongue and the resulting sound, the place of articulation is an important concept in many subdisciplines of phonetics.

Sounds are partly categorized by the location of a constriction as well as the part of the body doing the constricting. For example, in English the words *fought* and *thought* are a minimal pair differing only in the organ making the construction rather than the location of the construction.

The "f" in *fought* is a labiodental articulation made with the bottom lip against the teeth. The "th" in *thought* is a linguodental articulation made with the tongue against the teeth. Constrictions made by the lips are calledlabials while those made with the tongue are called lingual.

Constrictions made with the tongue can be made in several parts of the vocal tract, broadly classified into coronal, dorsal and radical places of articulation. Coronal articulations are made with the front of the tongue, dorsal articulations are made with the back of the tongue, and radical articulations are made in the pharynx.

These divisions are not sufficient for distinguishing and describing all speech sounds. For example, in English the sounds [s] and [ʃ] are both coronal, but they are produced in different places of the mouth. To account for this, more

detailed places of articulation are needed based upon the area of the mouth in which the constriction occurs.

Labial

Articulations involving the lips can be made in three different ways: with both lips (bilabial), with one lip and the teeth (labiodental), and with the tongue and the upper lip (linguolabial). Depending on the definition used, some or all of these kinds of articulations may be categorized into the class of labial articulations. Bilabial consonantsare made with both lips. In producing these sounds the lower lip moves farthest to meet the upper lip, which also moves down slightly, though in some cases the force from air moving through the aperture (opening between the lips) may cause the lips to separate faster than they can come together. Unlike most other articulations, both articulators are made from soft tissue, and so bilabial stops are more likely to be produced with incomplete closures than articulations involving hard surfaces like the teeth or palate. Bilabial stops are also unusual in that an articulator in the upper section of the vocal tract actively moves downwards, as the upper lip shows some active downward movement. Linguolabial consonantsare made with the blade of the tongue approaching or contacting the upper lip. Like in bilabial articulations, the upper lip moves slightly towards the more active articulator. Articulations in this group do not have their own symbols in the International Phonetic Alphabet, rather, they are formed by combining an apical symbol with a diacritic implicitly placing them in the coronal category. They exist in a number of languages indigenous to Vanuatu such as Tangoa.

Labiodental consonants are made by the lower lip rising to the upper teeth. Labiodental consonants are most often fricatives while labiodental nasals are also typologically common. There is debate as to whether true labiodental plosives occur in any natural language, though a number of languages are reported to have labiodental plosives including Zulu, Tonga, and Shubi.

Coronal

Coronal consonants are made with the tip or blade of the tongue and, because of the agility of the front of the tongue, represent a variety not only in place but in the posture of the tongue. The coronal places of articulation represent the areas of the mouth where the tongue contacts or makes a constriction, and include dental, alveolar, and post-alveolar locations. Tongue postures using the tip of the tongue can be apical if using the top of the tongue tip, laminal if made with the blade of the tongue, or sub-apical if the tongue tip is curled back and the bottom of the tongue is used. Coronals are unique as a group in that every manner of articulation is attested.

Australian languagesare well known for the large number of coronal contrasts exhibited within and across languages in the region. Dental consonantsare made with the tip or blade of the tongue and the upper teeth. They are divided into two groups based upon the part of the tongue used to produce them: apical dental consonants are produced with the tongue tip touching the teeth; interdental consonants are produced with the blade of the tongue as the tip of the tongue sticks out in front of the teeth. No language is known to use both contrastively though they may exist allophonically. Alveolar

consonantsare made with the tip or blade of the tongue at the alveolar ridge just behind the teeth and can similarly be apical or laminal.

Crosslinguistically, dental consonants and alveolar consonants are frequently contrasted leading to a number of generalizations of crosslinguistic patterns. The different places of articulation tend to also be contrasted in the part of the tongue used to produce them: most languages with dental stops have laminal dentals, while languages with apical stops usually have apical stops. Languages rarely have two consonants in the same place with a contrast in laminality, though Taa (!Xóõ) is a counterexample to this pattern. If a language has only one of a dental stop or an alveolar stop, it will usually be laminal if it is a dental stop, and the stop will usually be apical if it is an alveolar stop, though for example Temne and Bulgarian do not follow this pattern. If a language has both an apical and laminal stop, then the laminal stop is more likely to be affricated like in Isoko, though Dahalo show the opposite pattern with alveolar stops being more affricated.

Retroflex consonants have several different definitions depending on whether the position of the tongue or the position on the roof of the mouth is given prominence. In general, they represent a group of articulations in which the tip of the tongue is curled upwards to some degree. In this way, retroflex articulations can occur in several different locations on the roof of the mouth including alveolar, post-alveolar, and palatal regions.

If the underside of the tongue tip makes contact with the roof of the mouth, it is sub-apical though apical post-alveolar

sounds are also described as retroflex. Typical examples of sub-apical retroflex stops are commonly found in Dravidian languages, and in some languages indigenous to the southwest United States the contrastive difference between dental and alveolar stops is a slight retroflexion of the alveolar stop. Acoustically, retroflexion tends to affect the higher formants.

Articulations taking place just behind the alveolar ridge, known as post-alveolar consonants, have been referred to using a number of different terms. Apical post-alveolar consonants are often called retroflex, while laminal articulations are sometimes called palato-alveolar; in the Australianist literature, these laminal stops are often described as 'palatal' though they are produced further forward than the palate region typically described as palatal. Because of individual anatomical variation, the precise articulation of palato-alveolar stops (and coronals in general) can vary widely within a speech community.

Dorsal

Dorsal consonants are those consonants made using the tongue body rather than the tip or blade and are typically produced at the palate, velum or uvula. Palatal consonantsare made using the tongue body against the hard palate on the roof of the mouth. They are frequently contrasted with velar or uvular consonants, though it is rare for a language to contrast all three simultaneously, with Jaqaru as a possible example of a three-way contrast. Velar consonantsare made using the tongue body against the velum. They are incredibly common cross-linguistically; almost all languages have a velar stop. Because both velars and vowels are made using the tongue

body, they are highly affected by coarticulation with vowels and can be produced as far forward as the hard palate or as far back as the uvula. These variations are typically divided into front, central, and back velars in parallel with the vowel space. They can be hard to distinguish phonetically from palatal consonants, though are produced slightly behind the area of prototypical palatal consonants.

Uvular consonants are made by the tongue body contacting or approaching the uvula. They are rare, occurring in an estimated 19 percent of languages, and large regions of the Americas and Africa have no languages with uvular consonants. In languages with uvular consonants, stops are most frequent followed by continuants (including nasals).

Pharyngeal and laryngeal

Consonants made by constrictions of the throat are pharyngeals, and those made by a constriction in the larynx are laryngeal. Laryngeals are made using the vocal folds as the larynx is too far down the throat to reach with the tongue. Pharyngeals however are close enough to the mouth that parts of the tongue can reach them.

Radical consonants either use the root of the tongue or the epiglottis during production and are produced very far back in the vocal tract. Pharyngeal consonants are made by retracting the root of the tongue far enough to almost touch the wall of the pharynx. Due to production difficulties, only fricatives and approximants can produced this way. Epiglottal consonantsare made with the epiglottis and the back wall of the pharynx. Epiglottal stops have been recorded in Dahalo. Voiced

epiglottal consonants are not deemed possible due to the cavity between the glottis and epiglottis being too small to permit voicing.

Glottal consonants are those produced using the vocal folds in the larynx. Because the vocal folds are the source of phonation and below the oro-nasal vocal tract, a number of glottal consonants are impossible such as a voiced glottal stop. Three glottal consonants are possible, a voiceless glottal stop and two glottal fricatives, and all are attested in natural languages. Glottal stops, produced by closing the vocal folds, are notably common in the world's languages. While many languages use them to demarcate phrase boundaries, some languages like Huatla Mazatec have them as contrastive phonemes. Additionally, glottal stops can be realized as laryngealization of the following vowel in this language. Glottal stops, especially between vowels, do usually not form a complete closure. True glottal stops normally occur only when they'regeminated.

The larynx

The larynx, commonly known as the "voice box", is a cartilaginous structure in the trachea responsible for phonation. The vocal folds (chords) are held together so that they vibrate, or held apart so that they do not. The positions of the vocal folds are achieved by movement of the arytenoid cartilages. The intrinsic laryngeal muscles are responsible for moving the arytenoid cartilages as well as modulating the tension of the vocal folds. If the vocal folds are not close or tense enough, they will either vibrate sporadically or not at all. If they vibrate sporadically it will result in either creaky or

breathy voice, depending on the degree; if don't vibrate at all, the result will be voicelessness.

In addition to correctly positioning the vocal folds, there must also be air flowing across them or they will not vibrate. The difference in pressure across the glottis required for voicing is estimated at 1 – 2 cm H_2O (98.0665 – 196.133 pascals). The pressure differential can fall below levels required for phonation either because of an increase in pressure above the glottis (superglottal pressure) or a decrease in pressure below the glottis (subglottal pressure).

The subglottal pressure is maintained by the respiratory muscles. Supraglottal pressure, with no constrictions or articulations, is equal to about atmospheric pressure. However, because articulations—especially consonants—represent constrictions of the airflow, the pressure in the cavity behind those constrictions can increase resulting in a higher supraglottal pressure.

Lexical access

According to the lexical access model two different stages of cognition are employed; thus, this concept is known as the two-stage theory of lexical access. The first stage, lexical selection provides information about lexical items required to construct the functional level representation. These items are retrieved according to their specific semantic and syntactic properties, but phonological forms are not yet made available at this stage. The second stage, retrieval of wordforms, provides information required for building the positional level representation.

Articulatory models

When producing speech, the articulators move through and contact particular locations in space resulting in changes to the acoustic signal. Some models of speech production take this as the basis for modeling articulation in a coordinate system that may be internal to the body (intrinsic) or external (extrinsic). Intrinsic coordinate systems model the movement of articulators as positions and angles of joints in the body. Intrinsic coordinate models of the jaw often use two to three degrees of freedom representing translation and rotation. These face issues with modeling the tongue which, unlike joints of the jaw and arms, is a muscular hydrostat—like an elephant trunk—which lacks joints. Because of the different physiological structures, movement paths of the jaw are relatively straight lines during speech and mastication, while movements of the tongue follow curves.

Straight-line movements have been used to argue articulations as planned in extrinsic rather than intrinsic space, though extrinsic coordinate systems also include acoustic coordinate spaces, not just physical coordinate spaces. Models that assume movements are planned in extrinsic space run into an inverse problem of explaining the muscle and joint locations which produce the observed path or acoustic signal. The arm, for example, has seven degrees of freedom and 22 muscles, so multiple different joint and muscle configurations can lead to the same final position. For models of planning in extrinsic acoustic space, the same one-to-many mapping problem applies as well, with no unique mapping from physical or acoustic targets to the muscle movements required to achieve them. Concerns about the inverse problem may be exaggerated,

however, as speech is a highly learned skill using neurological structures which evolved for the purpose.

The equilibrium-point model proposes a resolution to the inverse problem by arguing that movement targets be represented as the position of the muscle pairs acting on a joint. Importantly, muscles are modeled as springs, and the target is the equilibrium point for the modeled spring-mass system. By using springs, the equilibrium point model can easily account for compensation and response when movements are disrupted. They are considered a coordinate model because they assume that these muscle positions are represented as points in space, equilibrium points, where the spring-like action of the muscles converges.

Gestural approaches to speech production propose that articulations are represented as movement patterns rather than particular coordinates to hit. The minimal unit is a gesture that represents a group of "functionally equivalent articulatory movement patterns that are actively controlled with reference to a given speech-relevant goal (e.g., a bilabial closure)." These groups represent coordinative structures or "synergies" which view movements not as individual muscle movements but as task-dependent groupings of muscles which work together as a single unit. This reduces the degrees of freedom in articulation planning, a problem especially in intrinsic coordinate models, which allows for any movement that achieves the speech goal, rather than encoding the particular movements in the abstract representation. Coarticulationis well described by gestural models as the articulations at faster speech rates can be explained as composites of the independent gestures at slower speech rates.

Acoustics

Speech sounds are created by the modification of an airstream which results in a sound wave. The modification is done by the articulators, with different places and manners of articulation producing different acoustic results. Because the posture of the vocal tract, not just the position of the tongue can affect the resulting sound, the manner of articulation is important for describing the speech sound.

The words *tack* and *sack* both begin with alveolar sounds in English, but differ in how far the tongue is from the alveolar ridge. This difference has large affects on the air stream and thus the sound that is produced. Similarly, the direction and source of the airstream can affect the sound. The most common airstream mechanism is pulmonic—using the lungs—but the glottis and tongue can also be used to produce airstreams.

Voicing and phonation types

A major distinction between speech sounds is whether they are voiced. Sounds are voiced when the vocal folds begin to vibrate in the process of phonation. Many sounds can be produced with or without phonation, though physical constraints may make phonation difficult or impossible for some articulations. When articulations are voiced, the main source of noise is the periodic vibration of the vocal folds. Articulations like voiceless plosives have no acoustic source and are noticeable by their silence, but other voiceless sounds like fricatives create their own acoustic source regardless of phonation.

Phonation is controlled by the muscles of the larynx, and languages make use of more acoustic detail than binary voicing. During phonation, the vocal folds vibrate at a certain rate. This vibration results in a periodic acoustic waveform comprising a fundamental frequency and its harmonics. The fundamental frequency of the acoustic wave can be controlled by adjusting the muscles of the larynx, and listeners perceive this fundamental frequency as pitch. Languages use pitch manipulation to convey lexical information in tonal languages, and many languages use pitch to mark prosodic or pragmatic information.

For the vocal folds to vibrate, they must be in the proper position and there must be air flowing through the glottis. Phonation types are modeled on a continuum of glottal states from completely open (voiceless) to completely closed (glottal stop). The optimal position for vibration, and the phonation type most used in speech, modal voice, exists in the middle of these two extremes. If the glottis is slightly wider, breathy voice occurs, while bringing the vocal folds closer together results in creaky voice.

The normal phonation pattern used in typical speech is modal voice, where the vocal folds are held close together with moderate tension. The vocal folds vibrate as a single unit periodically and efficiently with a full glottal closure and no aspiration. If they are pulled farther apart, they do not vibrate and so produce voiceless phones. If they are held firmly together they produce a glottal stop.

If the vocal folds are held slightly further apart than in modal voicing, they produce phonation types like breathy voice (or

murmur) and whispery voice. The tension across the vocal ligaments (vocal cords) is less than in modal voicing allowing for air to flow more freely. Both breathy voice and whispery voice exist on a continuum loosely characterized as going from the more periodic waveform of breathy voice to the more noisy waveform of whispery voice. Acoustically, both tend to dampen the first formant with whispery voice showing more extreme deviations. Holding the vocal folds more tightly together results in a creaky voice. The tension across the vocal folds is less than in modal voice, but they are heldtightly together resulting in only the ligaments of the vocal folds vibrating. The pulses are highly irregular, with low pitch and frequency amplitude.

Some languages do not maintain a voicing distinction for some consonants, but all languages use voicing to some degree. For example, no language is known to have a phonemic voicing contrast for vowels with all known vowels canonically voiced. Other positions of the glottis, such as breathy and creaky voice, are used in a number of languages, like Jalapa Mazatec, to contrast phonemes while in other languages, like English, they exist allophonically.

There are several ways to determine if a segment is voiced or not, the simplest being to feel the larynx during speech and note when vibrations are felt. More precise measurements can be obtained through acoustic analysis of a spectrogram or spectral slice.

In a spectrographic analysis, voiced segments show a voicing bar, a region of high acoustic energy, in the low frequencies of voiced segments. In examining a spectral splice, the acoustic spectrum at a given point in time a model of the vowel

pronounced reverses the filtering of the mouth producing the spectrum of the glottis. A computational model of the unfiltered glottal signal is then fitted to the inverse filtered acoustic signal to determine the characteristics of the glottis. Visual analysis is also available using specialized medical equipment such as ultrasound and endoscopy.

Vowels

Vowels are broadly categorized by the area of the mouth in which they are produced, but because they are produced without a constriction in the vocal tract their precise description relies on measuring acoustic correlates of tongue position. The location of the tongue during vowel production changes the frequencies at which the cavity resonates, and it is these resonances—known as formants—which are measured and used to characterize vowels.

Vowel height traditionally refers to the highest point of the tongue during articulation. The height parameter is divided into four primary levels: high (close), close-mid, open-mid and low (open). Vowels whose height are in the middle are referred to as mid. Slightly opened close vowels and slightly closed open vowels are referred to as near-close and near-open respectively. The lowest vowels are not just articulated with a lowered tongue, but also by lowering the jaw.

While the IPA implies that there are seven levels of vowel height, it is unlikely that a given language can minimally contrast all seven levels. Chomsky and Halle suggest that there are only three levels, although four levels of vowel height seem

to be needed to describe Danish and it's possible that some languages might even need five.

Vowel backness is dividing into three levels: front, central and back. Languages usually do not minimally contrast more than two levels of vowel backness. Some languages claimed to have a three-way backness distinction include Nimboran and Norwegian.

In most languages, the lips during vowel production can be classified as either rounded or unrounded (spread), although other types of lip positions, such as compression and protrusion, have been described. Lip position is correlated with height and backness: front and low vowels tend to be unrounded whereas back and high vowels are usually rounded. Paired vowels on the IPA chart have the spread vowel on the left and the rounded vowel on the right.

Together with the universal vowel features described above, some languages have additional features such as nasality, length and different types of phonation such as voiceless or creaky. Sometimes more specialized tongue gestures such as rhoticity, advanced tongue root, pharyngealization, stridency and frication are required to describe a certain vowel.

Manner of articulation

Knowing the place of articulation is not enough to fully describe a consonant, the way in which the stricture happens is equally important. Manners of articulation describe how exactly the active articulator modifies, narrows or closes off the vocal tract.

Stops (also referred to as plosives) are consonants where the airstream is completely obstructed. Pressure builds up in the mouth during the stricture, which is then released as a small burst of sound when the articulators move apart. The velum is raised so that air cannot flow through the nasal cavity. If the velum is lowered and allows for air to flow through the nose, the result in a nasal stop. However, phoneticians almost always refer to nasal stops as just "nasals".Affricates are a sequence of stops followed by a fricative in the same place.

Fricatives are consonants where the airstream is made turbulent by partially, but not completely, obstructing part of the vocal tract. Sibilants are a special type of fricative where the turbulent airstream is directed towards the teeth, creating a high-pitched hissing sound.

Nasals (sometimes referred to as nasal stops) are consonants in which there's a closure in the oral cavity and the velum is lowered, allowing air to flow through the nose.

In an approximant, the articulators come close together, but not to such an extent that allows a turbulent airstream.

Laterals are consonants in which the airstream is obstructed along the center of the vocal tract, allowing the airstream to flow freely on one or both sides. Laterals have also been defined as consonants in which the tongue is contracted in such a way that the airstream is greater around the sides than over the center of the tongue. The first definition does not allow for air to flow over the tongue.

Trills are consonants in which the tongue or lips are set in motion by the airstream. The stricture is formed in such a way

that the airstream causes a repeating pattern of opening and closing of the soft articulator(s). Apical trills typically consist of two or three periods of vibration.

Taps and flaps are single, rapid, usually apical gestures where the tongue is thrown against the roof of the mouth, comparable to a very rapid stop. These terms are sometimes used interchangeably, but some phoneticians make a distinction. In a tap, the tongue contacts the roof in a single motion whereas in a flap the tongue moves tangentially to the roof of the mouth, striking it in passing.

During a glottalic airstream mechanism, the glottis is closed, trapping a body of air. This allows for the remaining air in the vocal tract to be moved separately. An upward movement of the closed glottis will move this air out, resulting in it an ejective consonant. Alternatively, the glottis can lower, sucking more air into the mouth, which results in an implosive consonant.

Clicks are stops in which tongue movement causes air to be sucked in the mouth, this is referred to as a velaric airstream. During the click, the air becomes rarefied between two articulatory closures, producing a loud 'click' sound when the anterior closure is released. The release of the anterior closure is referred to as the click influx. The release of the posterior closure, which can be velar or uvular, is the click efflux. Clicks are used in several African language families, such as the Khoisan and Bantu languages.

Pulmonary and subglottal system

The lungs drive nearly all speech production, and their importance in phonetics is due to their creation of pressure for

pulmonic sounds. The most common kinds of sound across languages are pulmonic egress, where air is exhaled from the lungs. The opposite is possible, though no language is known to have pulmonic ingressive sounds as phonemes. Many languages such as Swedish use them for paralinguistic articulations such as affirmations in a number of genetically and geographically diverse languages. Both egressive and ingressive sounds rely on holding the vocal folds in a particular posture and using the lungs to draw air across the vocal folds so that they either vibrate (voiced) or do not vibrate (voiceless). Pulmonic articulations are restricted by the volume of air able to be exhaled in a given respiratory cycle, known as the vital capacity.

The lungs are used to maintain two kinds of pressure simultaneously in order to produce and modify phonation. To produce phonation at all, the lungs must maintain a pressure of 3–5 cm H_2O higher than the pressure above the glottis. However small and fast adjustments are made to the subglottal pressure to modify speech for suprasegmental features like stress. A number of thoracic muscles are used to make these adjustments. Because the lungs and thorax stretch during inhalation, the elastic forces of the lungs alone can produce pressure differentials sufficient for phonation at lung volumes above 50 percent of vital capacity. Above 50 percent of vital capacity, the respiratory muscles are used to "check" the elastic forces of the thorax to maintain a stable pressure differential. Below that volume, they are used to increase the subglottal pressure by actively exhaling air.

During speech, the respiratory cycle is modified to accommodate both linguistic and biological needs. Exhalation,

usually about 60 percent of the respiratory cycle at rest, is increased to about 90 percent of the respiratory cycle. Because metabolic needs are relatively stable, the total volume of air moved in most cases of speech remains about the same as quiet tidal breathing. Increases in speech intensity of 18 dB (a loud conversation) has relatively little impact on the volume of air moved. Because their respiratory systems are not as developed as adults, children tend to use a larger proportion of their vital capacity compared to adults, with more deep inhales.

Source–filter theory

The source–filter model of speech is a theory of speech production which explains the link between vocal tract posture and the acoustic consequences. Under this model, the vocal tract can be modeled as a noise source coupled onto an acoustic filter. The noise source in many cases is the larynx during the process of voicing, though other noise sources can be modeled in the same way. The shape of the supraglottal vocal tract acts as the filter, and different configurations of the articulators result in different acoustic patterns. These changes are predictable. The vocal tract can be modeled as a sequence of tubes, closed at one end, with varying diameters, and by using equations for acoustic resonance the acoustic effect of an articulatory posture can be derived. The process of inverse filtering uses this principle to analyze the source spectrum produced by the vocal folds during voicing. By taking the inverse of a predicted filter, the acoustic effect of the supraglottal vocal tract can be undone giving the acoustic spectrum produced by the vocal folds. This allows quantitative study of the various phonation types.

Perception

Language perception is the process by which a linguistic signal is decoded and understood by a listener. In order to perceive speech the continuous acoustic signal must be converted into discrete linguistic units such as phonemes, morphemes, and words. In order to correctly identify and categorize sounds, listeners prioritize certain aspects of the signal that can reliably distinguish between linguistic categories. While certain cues are prioritized over others, many aspects of the signal can contribute to perception. For example, though oral languages prioritize acoustic information, the McGurk effect shows that visual information is used to distinguish ambiguous information when the acoustic cues are unreliable.

While listeners can use a variety of information to segment the speech signal, the relationship between acoustic signal and category perception is not a perfect mapping. Because of coarticulation, noisy environments, and individual differences, there is a high degree of acoustic variability within categories. Known as the problem of **perceptual invariance**, listeners are able to reliably perceive categories despite the variability in acoustic instantiation. In order to do this, listeners rapidly accommodate to new speakers and will shift their boundaries between categories to match the acoustic distinctions their conversational partner is making.

Audition

Audition, the process of hearing sounds, is the first stage of perceiving speech. Articulators cause systematic changes in air

pressure which travel as sound waves to the listener's ear. The sound waves then hit the listener's ear drum causing it to vibrate. The vibration of the ear drum is transmitted by the ossicles—three small bones of the middle ear—to the cochlea. The cochlea is a spiral-shaped, fluid-filled tube divided lengthwise by the organ of Corti which contains the basilar membrane. The basilar membrane increases in thickness as it travels through the cochlea causing different frequencies to resonate at different locations. This tonotopic design allows for the ear to analyze sound in a manner similar to a Fourier transform.

The differential vibration of the basilar causes the hair cells within the organ of Corti to move. This causes depolarization of the hair cells and ultimately a conversion of the acoustic signal into a neuronal signal. While the hair cells do not produce action potentials themselves, they release neurotransmitter at synapses with the fibers of the auditory nerve, which does produce action potentials.

In this way, the patterns of oscillations on the basilar membrane are converted to spatiotemporal patterns of firings which transmit information about the sound to the brainstem.

Prosody

Besides consonants and vowels, phonetics also describes the properties of speech that are not localized to segments but greater units of speech, such as syllables and phrases. Prosody includes auditory characteristics such as pitch, speech rate, duration, and loudness.

Languages use these properties to different degrees to implement stress, pitch accents, and intonation — for example, stress in English and Spanishis correlated with changes in pitch and duration, whereas stress in Welsh is more consistently correlated with pitch than duration and stress in Thai is only correlated with duration.

Theories of speech perception

Early theories of speech perception such as motor theory attempted to solve the problem of perceptual invariance by arguing that speech perception and production were closely linked. In its strongest form, motor theory argues that speech perception *requires* the listener to access the articulatory representation of sounds; in order to properly categorize a sound, a listener reverse engineers the articulation which would produce that sound and by identifying these gestures is able to retrieve the intended linguistic category. While findings such as the McGurk effect and case studies from patients with neurological injuries have provided support for motor theory, further experiments have not supported the strong form of motor theory, though there is some support for weaker forms of motor theory which claim a non-deterministic relationship between production and perception.

Successor theories of speech perception place the focus on acoustic cues to sound categories and can be grouped into two broad categories: abstractionist theories and episodic theories. In abstractionist theories, speech perception involves the identification of an idealized lexical object based on a signal reduced to its necessary components and normalizing the signal to counteract speaker variability. Episodic theories such

as the exemplar model argue that speech perception involves accessing detailed memories (i.e., episodic memories) of previously heard tokens.

The problem of perceptual invariance is explained by episodic theories as an issue of familiarity: normalization is a byproduct of exposure to more variable distributions rather than a discrete process as abstractionist theories claim.

Subdisciplines

Acoustic phonetics

Acoustic phonetics deals with the acoustic properties of speech sounds. The sensation of sound is caused by pressure fluctuations which cause the eardrum to move. The ear transforms this movement into neural signals that the brain registers as sound. Acoustic waveforms are records that measure these pressure fluctuations.

Articulatory phonetics

Articulatory phonetics deals with the ways in which speech sounds are made.

Auditory phonetics

Auditory phonetics studies how humans perceive speech sounds. Due to the anatomical features of the auditory system distorting the speech signal, humans do not experience speech sounds as perfect acoustic records.

For example, the auditory impressions of volume, measured in decibels (dB), does not linearly match the difference in sound pressure.

The mismatch between acoustic analyses and what the listener hears is especially noticeable in speech sounds that have a lot of high-frequency energy, such as certain fricatives. To reconcile this mismatch, functional models of the auditory system have been developed.

Describing sounds

Human languages use many different sounds and in order to compare them linguists must be able to describe sounds in a way that is language independent. Speech sounds can be described in a number of ways. Most commonly speech sounds are referred to by the mouth movements needed to produce them. Consonants and vowels are two gross categories that phoneticians define by the movements in a speech sound. More fine-grained descriptors are parameters such as place of articulation.

Place of articulation, manner of articulation, and voicingare used to describe consonants and are the main divisions of the International Phonetic Alphabet consonant chart. Vowels are described by their height, backness, and rounding. Sign language are described using a similar but distinct set of parameters to describe signs: location, movement, hand shape, palm orientation, and non-manual features. In addition to articulatory descriptions, sounds used in oral languages can be described using their acoustics. Because the acoustics are a consequence of the articulation, both methods of description

are sufficient to distinguish sounds with the choice between systems dependent on the phonetic feature being investigated.

Consonants are speech sounds that are articulated with a complete or partial closure of the vocal tract. They are generally produced by the modification of an airstream exhaled from the lungs. The respiratory organs used to create and modify airflow are divided into three regions: the vocal tract (supralaryngeal), the larynx, and the subglottal system. The airstream can be either egressive (out of the vocal tract) or ingressive (into the vocal tract).

In pulmonic sounds, the airstream is produced by the lungs in the subglottal system and passes through the larynx and vocal tract. Glottalic sounds use an airstream created by movements of the larynx without airflow from the lungs. Click consonants are articulated through the rarefaction of air using the tongue, followed by releasing the forward closure of the tongue.

Vowels are syllabic speech sounds that are pronounced without any obstruction in the vocal tract. Unlike consonants, which usually have definite places of articulation, vowels are defined in relation to a set of reference vowels called cardinal vowels. Three properties are needed to define vowels: tongue height, tongue backness and lip roundedness. Vowels that are articulated with a stable quality are called monophthongs; a combination of two separate vowels in the same syllable is a diphthong. In the IPA, the vowels are represented on a trapezoid shape representing the human mouth: the vertical axis representing the mouth from floor to roof and the horizontal axis represents the front-back dimension.

Transcription

Phonetic transcription is a system for transcribing phones that occur in a language, whether oral or sign. The most widely known system of phonetic transcription, the International Phonetic Alphabet (IPA), provides a standardized set of symbols for oral phones. The standardized nature of the IPA enables its users to transcribe accurately and consistently the phones of different languages, dialects, and idiolects. The IPA is a useful tool not only for the study of phonetics, but also for language teaching, professional acting, and speech pathology.

While no sign language has a standardized writing system, linguists have developed their own notation systems that describe the handshape, location and movement.

The Hamburg Notation System (HamNoSys) is similar to the IPA in that it allows for varying levels of detail. Some notation systems such as KOMVA and the Stokoe systemwere designed for use in dictionaries; they also make use of alphabetic letters in the local language for handshapes whereas HamNoSys represents the handshape directly. SignWriting aims to be an easy-to-learn writing system for sign languages, although it has not been officially adopted by any deaf community yet.

Sign languages

Unlike spoken languages, words in sign languagesare perceived with the eyes instead of the ears. Signs are articulated with the hands, upper body and head. The main articulators are the hands and arms. Relative parts of the arm are described with the terms proximal and distal. Proximal refers to a part closer

to the torso whereas a distal part is further away from it. For example, a wrist movement is distal compared to an elbow movement. Due to requiring less energy, distal movements are generally easier to produce. Various factors – such as muscle flexibility or being consideredtaboo – restrict what can be considered a sign.

Native signers do not look at their conversation partner's hands. Instead, their gaze is fixated on the face. Because peripheral vision is not as focused as the center of the visual field, signs articulated near the face allow for more subtle differences in finger movement and location to be perceived.

Unlike spoken languages, sign languages have two identical articulators: the hands. Signers may use whichever hand they prefer with no disruption in communication. Due to universal neurological limitations, two-handed signs generally have the same kind of articulation in both hands; this is referred to as the Symmetry Condition.

The second universal constraint is the Dominance Condition, which holds that when two handshapes are involved, one hand will remain stationary and have a more limited set handshapes compared to the dominant, moving hand. Additionally, it is common for one hand in a two-handed sign to be dropped during informal conversations, a process referred to as weak drop. Just like words in spoken languages, coarticulation may cause signs to influence each other's form. Examples include the handshapes of neighboring signs becoming more similar to each other (assimilation) or weak drop (an instance of deletion).

Phonology

Phonology is a branch of linguistics that studies how languages or dialects systematically organize their sounds (or signs, in sign languages). The term also refers to the sound system of any particular language variety. At one time, the study of phonology only related to the study of the systems of phonemes in spoken languages. Now it may relate to

- (a) any linguistic analysis either at a level beneath the word (including syllable, onset and rime, articulatory gestures, articulatory features, mora, etc.), or
- (b) all levels of language where sound or signs are structured to convey linguistic meaning.

Sign languages have a phonological system equivalent to the system of sounds in spoken languages. The building blocks of signs are specifications for movement, location and handshape.

Terminology

The word 'phonology' (as in *the phonology of English*) can also refer to the phonological system (sound system) of a given language.

This is one of the fundamental systems which a language is considered to comprise, like its syntax, its morphology and its vocabulary.

Phonology is often distinguished from *phonetics*. While phonetics concerns the physical production, acoustic transmission and perception of the sounds of speech, phonology describes the way sounds function within a given language or across languages to encode meaning.

For many linguists, phonetics belongs to descriptive linguistics, and phonology to theoretical linguistics, although establishing the phonological system of a language is necessarily an application of theoretical principles to analysis of phonetic evidence. Note that this distinction was not always made, particularly before the development of the modern concept of the phoneme in the mid 20th century. Some subfields of modern phonology have a crossover with phonetics in descriptive disciplines such as psycholinguistics and speech perception, resulting in specific areas like articulatory phonology or laboratory phonology.

Derivation and definitions

The word *phonology* comes from Ancient Greekφωνή, *phōnḗ*, "voice, sound," and the suffix *-logy* (which is from Greek λόγος, *lógos*, "word, speech, subject of discussion"). Definitions of the term vary. Nikolai Trubetzkoy in *Grundzüge der Phonologie* (1939) defines phonology as "the study of sound pertaining to the system of language," as opposed to phonetics, which is "the study of sound pertaining to the act of speech" (the distinction between *language* and *speech* being basically Saussure's distinction between *langue* and *parole*). More recently,

Lass (1998) writes that phonology refers broadly to the subdiscipline of linguistics concerned with the sounds of

language, while in more narrow terms, "phonology proper is concerned with the function, behavior and organization of sounds as linguistic items." According to Clark *et al.* (2007), it means the systematic use of sound to encode meaning in any spoken human language, or the field of linguistics studying this use.

History

Early evidence for a systematic study of the sounds in a language appears in the 4th century BCE *Ashtadhyayi*, a Sanskrit grammar composed by Pāṇini. In particular the *Shiva Sutras*, an auxiliary text to the *Ashtadhyayi*, introduces what may be considered a list of the phonemes of the Sanskrit language, with a notational system for them that is used throughout the main text, which deals with matters of morphology, syntax and semantics.

Ibn Jinni of Mosul, a pioneer in phonology, wrote prolifically in the 10th century on Arabic morphology and phonology of Arabic in works such as *Kitāb Al-Munṣif, Kitāb Al-Muḥtasab,* *andKitāb Al-Khaṣāʾiṣ* [ar].

The study of phonology as it exists today is defined by the formative studies of the 19th-century Polish scholar Jan Baudouin de Courtenay, who (together with his students MikołajKruszewski and Lev Shcherba) shaped the modern usage of the term *phoneme* in a series of lectures in 1876–1877. The word *phoneme* had been coined a few years earlier in 1873 by the French linguist A. Dufriche-Desgenettes. In a paper read at 24 May meeting of the Société de Linguistique de Paris, Dufriche-Desgenettes proposed that *phoneme* serve as a

one-word equivalent for the German *Sprachlaut*. Baudouin de Courtenay's subsequent work, though often unacknowledged, is considered to be the starting point of modern phonology. He also worked on the theory of phonetic alternations (what is now called allophony and morphophonology), and may have had an influence on the work of Saussure according to E. F. K. Koerner.

An influential school of phonology in the interwar period was the Prague school. One of its leading members was Prince Nikolai Trubetzkoy, whose *Grundzüge der Phonologie* (*Principles of Phonology*), published posthumously in 1939, is among the most important works in the field from this period. Directly influenced by Baudouin de Courtenay, Trubetzkoy is considered the founder of morphophonology, although this concept had also been recognized by de Courtenay. Trubetzkoy also developed the concept of the *archiphoneme*. Another important figure in the Prague school was Roman Jakobson, who was one of the most prominent linguists of the 20th century. In 1968Noam Chomsky and Morris Halle published *The Sound Pattern of English* (SPE), the basis for generative phonology. In this view, phonological representations are sequences of segments made up of distinctive features. These features were an expansion of earlier work by Roman Jakobson, Gunnar Fant, and Morris Halle. The features describe aspects of articulation and perception, are from a universally fixed set, and have the binary values + or –. There are at least two levels of representation: underlying representation and surface phonetic representation. Ordered phonological rules govern how underlying representationis transformed into the actual pronunciation (the so-called surface form). An important consequence of the influence SPE

had on phonological theory was the downplaying of the syllable and the emphasis on segments. Furthermore, the generativists folded morphophonology into phonology, which both solved and created problems.

Natural phonology is a theory based on the publications of its proponent David Stampe in 1969 and (more explicitly) in 1979. In this view, phonology is based on a set of universal phonological processes that interact with one another; which ones are active and which are suppressed is language-specific. Rather than acting on segments, phonological processes act on distinctive features within prosodic groups. Prosodic groups can be as small as a part of a syllable or as large as an entire utterance. Phonological processes are unordered with respect to each other and apply simultaneously (though the output of one process may be the input to another). The second most prominent natural phonologist is Patricia Donegan (Stampe's wife); there are many natural phonologists in Europe, and a few in the U.S., such as Geoffrey Nathan. The principles of natural phonology were extended to morphology by Wolfgang U. Dressler, who founded natural morphology.

In 1976, John Goldsmith introduced autosegmental phonology. Phonological phenomena are no longer seen as operating on *one* linear sequence of segments, called phonemes or feature combinations, but rather as involving *some parallel sequences* of features which reside on multiple tiers. Autosegmental phonology later evolved into feature geometry, which became the standard theory of representation for theories of the organization of phonology as different as lexical phonology and optimality theory.

Government phonology, which originated in the early 1980s as an attempt to unify theoretical notions of syntactic and phonological structures, is based on the notion that all languages necessarily follow a small set of principles and vary according to their selection of certain binary parameters. That is, all languages' phonological structures are essentially the same, but there is restricted variation that accounts for differences in surface realizations. Principles are held to be inviolable, though parameters may sometimes come into conflict. Prominent figures in this field include Jonathan Kaye, Jean Lowenstamm, Jean-Roger Vergnaud, MonikCharette, and John Harris.

In a course at the LSA summer institute in 1991, Alan Prince and Paul Smolensky developed optimality theory—an overall architecture for phonology according to which languages choose a pronunciation of a word that best satisfies a list of constraints ordered by importance; a lower-ranked constraint can be violated when the violation is necessary in order to obey a higher-ranked constraint. The approach was soon extended to morphology by John McCarthy and Alan Prince, and has become a dominant trend in phonology. The appeal to phonetic grounding of constraints and representational elements (e.g. features) in various approaches has been criticized by proponents of 'substance-free phonology', especially by Mark Hale and Charles Reiss.

An integrated approach to phonological theory that combines synchronic and diachronic accounts to sound patterns was initiated with Evolutionary Phonology in recent years.

Analysis of phonemes

- An important part of traditional, pre-generative schools of phonology is studying which sounds can be grouped into distinctive units within a language; these units are known as phonemes. For example, in English, the "p" sound in *pot* is aspirated (pronounced [p^h]) while that in *spot* is not aspirated (pronounced [p]). However, English speakers intuitively treat both sounds as variations (allophones) of the same phonological category, that is of the phoneme /p/. (Traditionally, it would be argued that if an aspirated [p^h] were interchanged with the unaspirated [p] in *spot*, native speakers of English would still hear the same words; that is, the two sounds are perceived as "the same" /p/.) In some other languages, however, these two sounds are perceived as different, and they are consequently assigned to different phonemes. For example, in Thai, Bengali, and Quechua, there are minimal pairs of words for which aspiration is the only contrasting feature (two words can have different meanings but with the only difference in pronunciation being that one has an aspirated sound where the other has an unaspirated one).

Part of the phonological study of a language therefore involves looking at data (phonetic transcriptions of the speech of native speakers) and trying to deduce what the underlying phonemes

are and what the sound inventory of the language is. The presence or absence of minimal pairs, as mentioned above, is a frequently used criterion for deciding whether two sounds should be assigned to the same phoneme. However, other considerations often need to be taken into account as well.

The particular contrasts which are phonemic in a language can change over time. At one time, [f] and [v], two sounds that have the same place and manner of articulation and differ in voicing only, were allophones of the same phoneme in English, but later came to belong to separate phonemes. This is one of the main factors of historical change of languages as described in historical linguistics.

The findings and insights of speech perception and articulation research complicate the traditional and somewhat intuitive idea of interchangeable allophones being perceived as the same phoneme. First, interchanged allophones of the same phoneme can result in unrecognizable words. Second, actual speech, even at a word level, is highly co-articulated, so it is problematic to expect to be able to splice words into simple segments without affecting speech perception.

Different linguists therefore take different approaches to the problem of assigning sounds to phonemes. For example, they differ in the extent to which they require allophones to be phonetically similar. There are also differing ideas as to whether this grouping of sounds is purely a tool for linguistic analysis, or reflects an actual process in the way the human brain processes a language.

Since the early 1960s, theoretical linguists have moved away from the traditional concept of a phoneme, preferring to

consider basic units at a more abstract level, as a component of morphemes; these units can be called*morphophonemes*, and analysis using this approach is called morphophonology.

Other topics in phonology

In addition to the minimal units that can serve the purpose of differentiating meaning (the phonemes), phonology studies how sounds alternate, i.e. replace one another in different forms of the same morpheme (allomorphs), as well as, for example, syllable structure, stress, feature geometry, and intonation.

Phonology also includes topics such as phonotactics (the phonological constraints on what sounds can appear in what positions in a given language) and phonological alternation (how the pronunciation of a sound changes through the application of phonological rules, sometimes in a given order which can be feeding or bleeding,) as well as prosody, the study of suprasegmentals and topics such as stress and intonation.

The principles of phonological analysis can be applied independently of modality because they are designed to serve as general analytical tools, not language-specific ones. The same principles have been applied to the analysis of sign languages (see Phonemes in sign languages), even though the sub-lexical units are not instantiated as speech sounds.

Chapter 3

Morphology is the Study of Word Structures

In linguistics, morphology (/mɔːrˈfɒlədʒi/) is the study of words, how they are formed, and their relationship to other words in the same language. It analyzes the structure of words and parts of words such as stems, root words, prefixes, and suffixes. Morphology also looks at parts of speech, intonation and stress, and the ways context can change a word's pronunciation and meaning. Morphology differs from morphological typology, which is the classification of languages based on their use of words, and lexicology, which is the study of words and how they make up a language's vocabulary.

While words, along with clitics, are generally accepted as being the smallest units of syntax, in most languages, if not all, many words can be related to other words by rules that collectively describe the grammar for that language. For example, English speakers recognize that the words *dog* and *dogs* are closely related, differentiated only by the pluralitymorpheme "-s", only found bound to noun phrases. Speakers of English, a fusional language, recognize these relations from their innate knowledge of English's rules of word formation. They infer intuitively that *dog* is to *dogs* as *cat* is to *cats*; and, in similar fashion, *dog* is to *dog catcher* as *dish* is to *dishwasher*. By contrast, Classical Chinese has very little morphology, using almost exclusively unbound morphemes ("free" morphemes) and depending on word order to convey meaning. (Most words in modern Standard Chinese

["Mandarin"], however, are compounds and most roots are bound.) These are understood as grammars that represent the morphology of the language. The rules understood by a speaker reflect specific patterns or regularities in the way words are formed from smaller units in the language they are using, and how those smaller units interact in speech. In this way, morphology is the branch of linguistics that studies patterns of word formation within and across languages and attempts to formulate rules that model the knowledge of the speakers of those languages.

Phonological and orthographic modifications between a base word and its origin may be partial to literacy skills. Studies have indicated that the presence of modification in phonology and orthography makes morphologically complex words harder to understand and that the absence of modification between a base word and its origin makes morphologically complex words easier to understand. Morphologically complex words are easier to comprehend when they include a base word.

Polysynthetic languages, such as Chukchi, have words composed of many morphemes. For example, the Chukchi word "təmeyŋəlevtpəɣtərkən", meaning "I have a fierce headache", is composed of eight morphemes *t-ə-meyŋ-ə-levt-pəɣt-ə-rkən* that may be glossed. The morphology of such languages allows for each consonant and vowel to be understood as morphemes, while the grammar of the language indicates the usage and understanding of each morpheme.

The discipline that deals specifically with the sound changes occurring within morphemes is morphophonology.

History

The history of morphological analysis dates back to the ancient Indian linguist Pāṇini, who formulated the 3,959 rules of Sanskrit morphology in the text *Aṣṭādhyāyī* by using a constituency grammar. The Greco-Roman grammatical tradition also engaged in morphological analysis. Studies in Arabic morphology, conducted by Marāḥ al-arwāḥ and Aḥmad b. ʻalīMasʻūd, date back to at least 1200 CE.

The linguistic term "morphology" was coined by August Schleicher in 1859.

Fundamental concepts

Lexemes and word forms

The term "word" has no well-defined meaning. Instead, two related terms are used in morphology: lexeme and word-form. Generally, a lexeme is a set of inflected word-forms that is often represented with the citation form in small capitals. For instance, the lexeme eat contains the word-forms *eat*, *eats*, *eaten*, and *ate*. *Eat* and *eats*are thus considered different word-forms belonging to the same lexeme eat. *Eat* and *Eater*, on the other hand, are different lexemes, as they refer to two different concepts.

Prosodic word vs. morphological word

Here are examples from other languages of the failure of a single phonological word to coincide with a single

morphological word form. In Latin, one way to express the concept of 'NOUN-PHRASE$_1$ and NOUN-PHRASE$_2$' (as in "apples and oranges") is to suffix '-que' to the second noun phrase: "apples oranges-and", as it were. An extreme level of this theoretical quandary posed by some phonological words is provided by the Kwak'wala language.

In Kwak'wala, as in a great many other languages, meaning relations between nouns, including possession and "semantic case", are formulated by affixes instead of by independent "words". The three-word English phrase, "with his club", where 'with' identifies its dependent noun phrase as an instrument and 'his' denotes a possession relation, would consist of two words or even just one word in many languages.

Unlike most languages, Kwak'wala semantic affixes phonologically attach not to the lexeme they pertain to semantically, but to the preceding lexeme. Consider the following example (in Kwak'wala, sentences begin with what corresponds to an English verb):

kwixʔid-i-dabəgwanəma$_i$-χ-a q'asa-s-is$_i$t'alwagwayu

Morpheme by morpheme translation:

- kwixʔid-i-da = clubbed-PIVOT-DETERMINER
- bəgwanəma-χ-a = man-ACCUSATIVE-DETERMINER
- q'asa-s-is = otter-INSTRUMENTAL-3SG-POSSESSIVE
- t'alwagwayu = club
- "the man clubbed the otter with his club."

(Notation notes:

- accusative case marks an entity that something is done to.
- determiners are words such as "the", "this", "that".
- the concept of "pivot" is a theoretical construct that is not relevant to this discussion.)

That is, to the speaker of Kwak'wala, the sentence does not contain the "words" 'him-the-otter' or 'with-his-club' Instead, the markers *-i-da* (PIVOT-'the'), referring to "man", attaches not to the noun *bəgwanəma* ("man") but to the verb; the markers *-χ-a* (ACCUSATIVE-'the'), referring to *otter*, attach to *bəgwanəma* instead of to *q'asa* ('otter'), etc. In other words, a speaker of Kwak'wala does not perceive the sentence to consist of these phonological words:

- kwix?idi-da-bəgwanəmaχ-a-q'asa s-is$_i$-t'alwagwayu
- clubbed PIVOT-the-man$_i$ hit-the-otter with-his$_i$-club

A central publication on this topic is the volume edited by Dixon and Aikhenvald (2002), examining the mismatch between prosodic-phonological and grammatical definitions of "word" in various Amazonian, Australian Aboriginal, Caucasian, Eskimo, Indo-European, Native North American, West African, and sign languages. Apparently, a wide variety of languages make use of the hybrid linguistic unit clitic, possessing the grammatical features of independent words but the prosodic-phonological lack of freedom of bound morphemes. The intermediate status of clitics poses a considerable challenge to linguistic theory.

Inflection vs. word formation

Given the notion of a lexeme, it is possible to distinguish two kinds of morphological rules. Some morphological rules relate to different forms of the same lexeme; while other rules relate to different lexemes. Rules of the first kind are inflectional rules, while those of the second kind are rules of word formation.

The generation of the English plural *dogs* from *dog* is an inflectional rule, while compound phrases and words like *dog catcher* or *dishwasher* are examples of word formation. Informally, word formation rules form "new" words (more accurately, new lexemes), while inflection rules yield variant forms of the "same" word (lexeme).

The distinction between inflection and word formation is not at all clear cut. There are many examples where linguists fail to agree whether a given rule is inflection or word formation. The next section will attempt to clarify this distinction.

Word formation is a process where one combines two complete words, whereas with inflection you can combine a suffix with some verb to change its form to subject of the sentence. For example: in the present indefinite, we use ‘go’ with subject I/we/you/they and plural nouns, whereas for third person singular pronouns (he/she/it) and singular nouns we use ‘goes’. So this ‘-es’ is an inflectional marker and is used to match with its subject. A further difference is that in word formation, the resultant word may differ from its source word's grammatical category whereas in the process of inflection the word never changes its grammatical category.

Types of word formation

There is a further distinction between two primary kinds of morphological word formation: derivation and compounding. Compounding is a process of word formation that involves combining complete word forms into a single compound form. *Dog catcher*, therefore, is a compound, as both *dog* and *catcher* are complete word forms in their own right but are subsequently treated as parts of one form. Derivation involves affixing bound (i.e. non-independent) forms to existing lexemes, whereby the addition of the affix derives a new lexeme. The word *independent*, for example, is derived from the word *dependent* by using the prefix *in-*, while *dependent* itself is derived from the verb *depend*. There is also word formation in the processes of clipping in which a portion of a word is removed to create a new one, blending in which two parts of different words are blended into one, acronyms in which each letter of the new word represents a specific word in the representation i.e. NATO for North Atlantic Treaty Organization, borrowing in which words from one language are taken and used in another, and finally coinage in which a new word is created to represent a new object or concept.

Paradigms and morphosyntax

A linguistic paradigm is the complete set of related word forms associated with a given lexeme. The familiar examples of paradigms are the conjugations of verbs and the declensions of nouns. Also, arranging the word forms of a lexeme into tables, by classifying them according to shared inflectional categories such as tense, aspect, mood, number, gender or case, organizes such. For example, the personal pronouns in English

can be organized into tables, using the categories of person (first, second, third); number (singular vs. plural); gender (masculine, feminine, neuter); and case (nominative, oblique, genitive).

The inflectional categories used to group word forms into paradigms cannot be chosen arbitrarily; they must be categories that are relevant to stating the syntactic rules of the language. Person and number are categories that can be used to define paradigms in English, because English has grammatical agreement rules that require the verb in a sentence to appear in an inflectional form that matches the person and number of the subject. Therefore, the syntactic rules of English care about the difference between *dog* and *dogs*, because the choice between these two forms determines which form of the verb is used. However, no syntactic rule for the difference between *dog* and *dog catcher*, or *dependent* and *independent*. The first two are nouns and the second two are adjectives.

An important difference between inflection and word formation is that inflected word forms of lexemes are organized into paradigms that are defined by the requirements of syntactic rules, and there are no corresponding syntactic rules for word formation. The relationship between syntax and morphology is called "morphosyntax" and concerns itself with inflection and paradigms, not with word formation or compounding.

Allomorphy

Above, morphological rules are described as analogies between word forms: *dog* is to *dogs* as *cat* is to *cats* and as *dish* is to

dishes. In this case, the analogy applies both to the form of the words and to their meaning: in each pair, the first word means "one of X", while the second "two or more of X", and the difference is always the plural form *-s* (or *-es*) affixed to the second word, signaling the key distinction between singular and plural entities.

One of the largest sources of complexity in morphology is that this one-to-one correspondence between meaning and form scarcely applies to every case in the language. In English, there are word form pairs like *ox*/*oxen*, *goose*/*geese*, and *sheep*/*sheep*, where the difference between the singular and the plural is signaled in a way that departs from the regular pattern, or is not signaled at all. Even cases regarded as regular, such as *-s*, are not so simple; the *-s* in *dogs* is not pronounced the same way as the *-s* in *cats*; and, in plurals such as *dishes*, a vowel is added before the *-s*. These cases, where the same distinction is effected by alternative forms of a "word", constitute allomorphy.

Phonological rules constrain which sounds can appear next to each other in a language, and morphological rules, when applied blindly, would often violate phonological rules, by resulting in sound sequences that are prohibited in the language in question. For example, to form the plural of *dish* by simply appending an *-s* to the end of the word would result in the form *[dɪʃs], which is not permitted by the phonotactics of English. In order to "rescue" the word, a vowel sound is inserted between the root and the plural marker, and [dɪʃɪz] results. Similar rules apply to the pronunciation of the *-s* in *dogs* and *cats*: it depends on the quality (voiced vs. unvoiced) of the final preceding phoneme.

Lexical morphology

Lexical morphology is the branch of morphology that deals with the lexicon, which, morphologically conceived, is the collection of lexemes in a language. As such, it concerns itself primarily with word formation: derivation and compounding.

Models

There are three principal approaches to morphology and each tries to capture the distinctions above in different ways:

- Morpheme-based morphology, which makes use of an item-and-arrangement approach.
- Lexeme-based morphology, which normally makes use of an item-and-process approach.
- Word-based morphology, which normally makes use of a word-and-paradigm approach.

While the associations indicated between the concepts in each item in that list are very strong, they are not absolute.

Morpheme-based morphology

In morpheme-based morphology, word forms are analyzed as arrangements of morphemes. A morpheme is defined as the minimal meaningful unit of a language. In a word such as *independently*, the morphemes are said to be *in-*, *de-*, *pend*, *-ent*, and *-ly*; *pend* is the (bound) root and the other morphemes are, in this case, derivational affixes. In words such as *dogs*, *dog* is the root and the *-s* is an inflectional morpheme. In its simplest and most naïve form, this way of analyzing word

forms, called "item-and-arrangement", treats words as if they were made of morphemes put after each other ("concatenated") like beads on a string. More recent and sophisticated approaches, such as distributed morphology, seek to maintain the idea of the morpheme while accommodating non-concatenated, analogical, and other processes that have proven problematic for item-and-arrangement theories and similar approaches. Morpheme-based morphology presumes three basic axioms:

- Baudouin's "single morpheme" hypothesis: Roots and affixes have the same status as morphemes.
- Bloomfield's "sign base" morpheme hypothesis: As morphemes, they are dualistic signs, since they have both (phonological) form and meaning.
- Bloomfield's "lexical morpheme" hypothesis: morphemes, affixes and roots alike are stored in the lexicon.

Morpheme-based morphology comes in two flavours, one Bloomfieldian and one Hockettian. For Bloomfield, the morpheme was the minimal form with meaning, but did not have meaning itself. For Hockett, morphemes are "meaning elements", not "form elements". For him, there is a morpheme plural using allomorphs such as *-s*, *-en* and *-ren*. Within much morpheme-based morphological theory, the two views are mixed in unsystematic ways so a writer may refer to "the morpheme plural" and "the morpheme *-s*" in the same sentence.

Lexeme-based morphology

Lexeme-based morphology usually takes what is called an item-and-process approach. Instead of analyzing a word form as a set of morphemes arranged in sequence, a word form is said to be the result of applying rules that alter a word-form or stem in order to produce a new one. An inflectional rule takes a stem, changes it as is required by the rule, and outputs a word form; a derivational rule takes a stem, changes it as per its own requirements, and outputs a derived stem; a compounding rule takes word forms, and similarly outputs a compound stem.

Word-based morphology

Word-based morphology is (usually) a word-and-paradigm approach. The theory takes paradigms as a central notion. Instead of stating rules to combine morphemes into word forms or to generate word forms from stems, word-based morphology states generalizations that hold between the forms of inflectional paradigms. The major point behind this approach is that many such generalizations are hard to state with either of the other approaches. Word-and-paradigm approaches are also well-suited to capturing purely morphological phenomena, such as morphomes. Examples to show the effectiveness of word-based approaches are usually drawn from fusional languages, where a given "piece" of a word, which a morpheme-based theory would call an inflectional morpheme, corresponds to a combination of grammatical categories, for example, "third-person plural". Morpheme-based theories usually have no problems with this situation since one says that a given morpheme has two categories. Item-and-process theories, on

the other hand, often break down in cases like these because they all too often assume that there will be two separate rules here, one for third person, and the other for plural, but the distinction between them turns out to be artificial. The approaches treat these as whole words that are related to each other by analogical rules. Words can be categorized based on the pattern they fit into. This applies both to existing words and to new ones. Application of a pattern different from the one that has been used historically can give rise to a new word, such as *older* replacing *elder* (where *older* follows the normal pattern of adjectivalsuperlatives) and *cows* replacing *kine* (where *cows* fits the regular pattern of plural formation).

Morphological typology

In the 19th century, philologists devised a now classic classification of languages according to their morphology. Some languages are isolating, and have little to no morphology; others are agglutinative whose words tend to have many easily separable morphemes; others yet are inflectional or fusional because their inflectional morphemes are "fused" together. That leads to one bound morpheme conveying multiple pieces of information. A standard example of an isolating language is Chinese. An agglutinative language is Turkish. Latin and Greek are prototypical inflectional or fusional languages.

It is clear that this classification is not at all clearcut, and many languages (Latin and Greek among them) do not neatly fit any one of these types, and some fit in more than one way. A continuum of complex morphology of language may be adopted.

The three models of morphology stem from attempts to analyze languages that more or less match different categories in this typology. The item-and-arrangement approach fits very naturally with agglutinative languages. The item-and-process and word-and-paradigm approaches usually address fusional languages.

As there is very little fusion involved in word formation, classical typology mostly applies to inflectional morphology. Depending on the preferred way of expressing non-inflectional notions, languages may be classified as synthetic (using word formation) or analytic (using syntactic phrases).

Examples

Pingelapese is a Micronesian language spoken on the Pingelap atoll and on two of the eastern Caroline Islands, called the high island of Pohnpei. Similar to other languages, words in Pingelapese can take different forms to add to or even change its meaning. Verbal suffixes are morphemes added at the end of a word to change its form. Prefixes are those that are added at the front. For example, the Pingelapese suffix *–kin* means 'with' or 'at.' It is added at the end of a verb.

ius = to use →*ius-kin* = to use with

mwahu = to be good →*mwahu-kin* = to be good at

sa- is an example of a verbal prefix. It is added to the beginning of a word and means 'not.'

pwung = to be correct →*sa-pwung* = to be incorrect

There are also directional suffixes that when added to the root word give the listener a better idea of where the subject is headed. The verb *alu* means to walk. A directional suffix can be used to give more detail.

-da = 'up' →*aluh-da* = to walk up

*-d*i = 'down' →*aluh-di* = to walk down

-eng = 'away from speaker and listener' →*aluh-eng* = to walk away

Directional suffixes are not limited to motion verbs. When added to non-motion verbs, their meanings are a figurative one. The following table gives some examples of directional suffixes and their possible meanings.

Chapter 4

Syntax is the Study of how Words are Combined to Form Sentences

In linguistics, syntax (/ˈsɪntæks/) is the study of how words and morphemes combine to form larger units such as phrases and sentences. Central concerns of syntax include word order, grammatical relations, hierarchical sentence structure (constituency), agreement, the nature of crosslinguistic variation, and the relationship between form and meaning. There are numerous approaches to syntax which differ in their central assumptions and goals.

Etymology

The word *syntax* comes from Ancient Greek: σύνταξις "coordination", which consists of σύν*syn*, "together", and τάξις*táxis*, "an ordering".

Sequencing of Subject, Verb, and Object

One basic description of a language's syntax is the sequence in which the subject (S), verb (V), and object (O) usually appear in sentences. Over 85% of languages usually place the subject first, either in the sequence SVO or the sequence SOV. The other possible sequences are VSO, VOS, OVS, and OSV, the last three of which are rare. In most generative theories of

syntax, these surface differences arise from a more complex clausal phrase structure, and each order may be compatible with multiple derivations.

Early history

The *Aṣṭādhyāyī* of Pāṇini (c. 4th century BC in Ancient India), is often cited as an example of a premodern work that approaches the sophistication of a modern syntactic theory (as works on grammar were written long before modern syntax came about). In the West, the school of thought that came to be known as "traditional grammar" began with the work of Dionysius Thrax.

For centuries, a framework known as *grammaire générale* (first expounded in 1660 by Antoine Arnauld in a book of the same title) dominated work in syntax: as its basic premise the assumption that language is a direct reflection of thought processes and therefore there is a single, most natural way to express a thought.

However, in the 19th century, with the development of historical-comparative linguistics, linguists began to realize the sheer diversity of human language and to question fundamental assumptions about the relationship between language and logic. It became apparent that there was no such thing as the most natural way to express a thought, and therefore logiccould no longer be relied upon as a basis for studying the structure of language.

The Port-Royal grammar modeled the study of syntax upon that of logic. (Indeed, large parts of the Port-Royal Logicwere copied or adapted from the *Grammairegénérale*.) Syntactic categories

were identified with logical ones, and all sentences were analyzed in terms of "subject – copula – predicate". Initially, this view was adopted even by the early comparative linguists such as Franz Bopp.

The central role of syntax within theoretical linguistics became clear only in the 20th century, which could reasonably be called the "century of syntactic theory" as far as linguistics is concerned. (For a detailed and critical survey of the history of syntax in the last two centuries, see the monumental work by Giorgio Graffi (2001).)

Theories of syntax

There are a number of theoretical approaches to the discipline of syntax. One school of thought, founded in the works of Derek Bickerton, sees syntax as a branch of biology, since it conceives of syntax as the study of linguistic knowledge as embodied in the human mind. Other linguists (e.g., Gerald Gazdar) take a more Platonistic view, since they regard syntax to be the study of an abstract formal system. Yet others (e.g., Joseph Greenberg) consider syntax a taxonomical device to reach broad generalizations across languages.

Syntacticians have attempted to explain the causes of word-order variation within individual languages and cross-linguistically. Much of such work has been done within frameworks of generative grammar which assumes that the core of syntax depends on a genetic structure which is common to all mankind. Typological research of the languages of the world has however found few absolute universals, leading some to conclude that none of syntax has to be directly genetic.

Alternative explanations have been sought in language processing. It is suggested that the brain finds it easier to parsesyntactic patterns which are either right or left branching, but not mixed. The most widely held approach is the performance–grammar correspondence hypothesis by John A. Hawkins who suggests that language is a non-innate adaptation to innate cognitive mechanisms. Cross-linguistic tendencies are considered as being based on language users' preference for grammars that are organized efficiently, and on their avoidance of word orderings which cause processing difficulty. Some languages however exhibit regular inefficient patterning. These include the VO languages Chinese, with the adpositional phrase before the verb, and Finnish which has postpositions; but there are few other profoundly exceptional languages.

Syntactic models

Dependency grammar

Dependency grammar (DG) is a class of modern grammatical theories that are all based on the dependency relation (as opposed to the *constituency relation* of phrase structure) and that can be traced back primarily to the work of Lucien Tesnière. Dependency is the notion that linguistic units, e.g. words, are connected to each other by directed links. The (finite) verb is taken to be the structural center of clause structure. All other syntactic units (words) are either directly or indirectly connected to the verb in terms of the directed links, which are called *dependencies*. Dependency grammar differs from phrase structure grammar in that while it can

identify phrases it tends to overlook phrasal nodes. A dependency structure is determined by the relation between a word (a head) and its dependents. Dependency structures are flatter than phrase structures in part because they lack a finiteverb phraseconstituent, and they are thus well suited for the analysis of languages with free word order, such as Czech or Warlpiri.

History

The notion of dependencies between grammatical units has existed since the earliest recorded grammars, e.g. Pāṇini, and the dependency concept therefore arguably predates that of phrase structure by many centuries. Ibn Maḍāʾ, a 12th-century linguist from Córdoba, Andalusia, may have been the first grammarian to use the term *dependency* in the grammatical sense that we use it today. In early modern times, the dependency concept seems to have coexisted side by side with that of phrase structure, the latter having entered Latin, French, English and other grammars from the widespread study of term logic of antiquity. Dependency is also concretely present in the works of SámuelBrassai (1800–1897), a Hungarian linguist, Franz Kern (1830-1894), a German philologist, and of HeimannHaritonTiktin (1850–1936), a Romanian linguist.

Modern dependency grammars, however, begin primarily with the work of Lucien Tesnière. Tesnière was a Frenchman, a polyglot, and a professor of linguistics at the universities in Strasbourg and Montpellier. His major work *Éléments de syntaxestructurale*was published posthumously in 1959 – he

died in 1954. The basic approach to syntax he developed seems to have been seized upon independently by others in the 1960s and a number of other dependency-based grammars have gained prominence since those early works. DG has generated a lot of interest in Germany in both theoretical syntax and language pedagogy. In recent years, the great development surrounding dependency-based theories has come from computational linguistics and is due, in part, to the influential work that David Hays did in machine translation at the RAND Corporation in the 1950s and 1960s. Dependency-based systems are increasingly being used to parse natural language and generate tree banks. Interest in dependency grammar is growing at present, international conferences on dependency linguistics being a relatively recent development (Depling 2011, Depling 2013, Depling 2015, Depling 2017, Depling 2019).

Dependency vs. phrase structure

- Dependency is a one-to-one correspondence: for every element (e.g. word or morph) in the sentence, there is exactly one node in the structure of that sentence that corresponds to that element. The result of this one-to-one correspondence is that dependency grammars are word (or morph) grammars. All that exist are the elements and the dependencies that connect the elements into a structure. This situation should be compared with phrase structure. Phrase structure is a one-to-one-or-more correspondence, which means that, for every element in a sentence, there is one or

more nodes in the structure that correspond to that element.

These trees illustrate two possible ways to render the dependency and phrase structure relations (see below). This dependency tree is an "ordered" tree, i.e. it reflects actual word order. Many dependency trees abstract away from linear order and focus just on hierarchical order, which means they do not show actual word order. This constituency (= phrase structure) tree follows the conventions of bare phrase structure (BPS), whereby the words themselves are employed as the node labels.

The distinction between dependency and phrase structure grammars derives in large part from the initial division of the clause. The phrase structure relation derives from an initial binary division, whereby the clause is split into a subject noun phrase (NP) and a predicateverb phrase (VP).

This division is certainly present in the basic analysis of the clause that we find in the works of, for instance, Leonard Bloomfield and Noam Chomsky. Tesnière, however, argued vehemently against this binary division, preferring instead to position the verb as the root of all clause structure. Tesnière's stance was that the subject-predicate division stems from term logic and has no place in linguistics. The importance of this distinction is that if one acknowledges the initial subject-predicate division in syntax is real, then one is likely to go down the path of phrase structure grammar, while if one rejects this division, then one must consider the verb as the root of all structure, and so go down the path of dependency grammar.

Dependency grammars

- The following frameworks are dependency-based:
- Algebraic syntax
- Operator grammar
- Link grammar
- Functional generative description
- Lexicase
- Meaning–text theory
- Word grammar
- Extensible dependency grammar
- Universal Dependencies

Link grammar is similar to dependency grammar, but link grammar does not include directionality between the linked words, and thus does not describe head-dependent relationships. Hybrid dependency/phrase structure grammar uses dependencies between words, but also includes dependencies between phrasal nodes – see for example the Quranic Arabic Dependency Treebank. The derivation trees of tree-adjoining grammar are dependency structures, although the full trees of TAG rendered in terms of phrase structure, so in this regard, it is not clear whether TAG should be viewed more as a dependency or phrase structure grammar.

There are major differences between the grammars just listed. In this regard, the dependency relation is compatible with other major tenets of theories of grammar. Thus like phrase structure grammars, dependency grammars can be mono- or

multistratal, representational or derivational, construction- or rule-based.

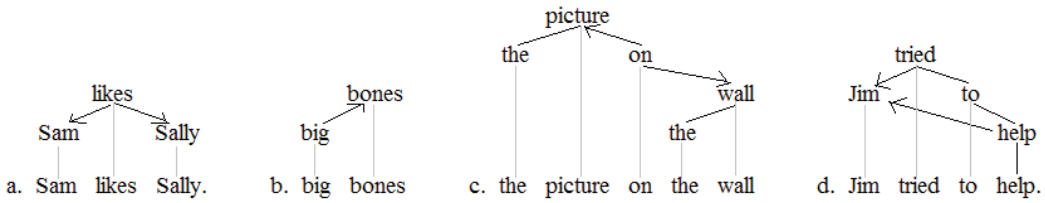
Types of dependencies

The dependency representations above (and further below) show syntactic dependencies. Indeed, most work in dependency grammar focuses on syntactic dependencies. Syntactic dependencies are, however, just one of three or four types of dependencies. Meaning–text theory, for instance, emphasizes the role of semantic and morphological dependencies in addition to syntactic dependencies. A fourth type, prosodic dependencies, can also be acknowledged. Distinguishing between these types of dependencies can be important, in part because if one fails to do so, the likelihood that semantic, morphological, and/or prosodic dependencies will be mistaken for syntactic dependencies is great. The following four subsections briefly sketch each of these dependency types. During the discussion, the existence of syntactic dependencies is taken for granted and used as an orientation point for establishing the nature of the other three dependency types.

Semantic dependencies

Semantic dependencies are understood in terms of predicates and their arguments. The arguments of a predicate are semantically dependent on that predicate. Often, semantic dependencies overlap with and point in the same direction as syntactic dependencies. At times, however, semantic dependencies can point in the opposite direction of syntactic dependencies, or they can be entirely independent of syntactic

dependencies. The hierarchy of words in the following examples show standard syntactic dependencies, whereas the arrows indicate semantic dependencies:

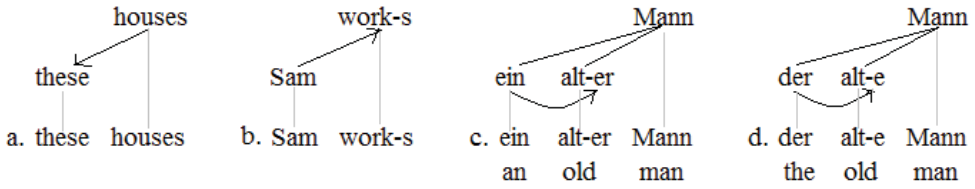


The two arguments *Sam* and *Sally* in tree (a) are dependent on the predicate *likes*, whereby these arguments are also syntactically dependent on *likes*. What this means is that the semantic and syntactic dependencies overlap and point in the same direction (down the tree). Attributive adjectives, however, are predicates that take their head noun as their argument, hence *big* is a predicate in tree (b) that takes *bones* as its one argument; the semantic dependency points up the tree and therefore runs counter to the syntactic dependency. A similar situation obtains in (c), where the preposition predicate *on* takes the two arguments *the picture* and *the wall*; one of these semantic dependencies points up the syntactic hierarchy, whereas the other points down it. Finally, the predicate *to help* in (d) takes the one argument *Jim* but is not directly connected to *Jim* in the syntactic hierarchy, which means that semantic dependency is entirely independent of the syntactic dependencies.

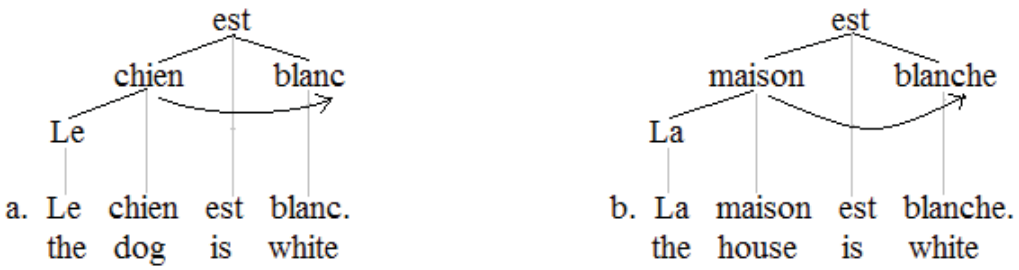
Morphological dependencies

Morphological dependencies obtain between words or parts of words. When a given word or part of a word influences the form of another word, then the latter is morphologically dependent

on the former. Agreement and concord are therefore manifestations of morphological dependencies. Like semantic dependencies, morphological dependencies can overlap with and point in the same direction as syntactic dependencies, overlap with and point in the opposite direction of syntactic dependencies, or be entirely independent of syntactic dependencies. The arrows are now used to indicate morphological dependencies.



The plural *houses* in (a) demands the plural of the demonstrative determiner, hence *these* appears, not *this*, which means there is a morphological dependency that points down the hierarchy from *houses* to *these*. The situation is reversed in (b), where the singular subject *Sam* demands the appearance of the agreement suffix *-s* on the finite verb *works*, which means there is a morphological dependency pointing up the hierarchy from *Sam* to *works*. The type of determiner in the German examples (c) and (d) influences the inflectional suffix that appears on the adjective *alt*. When the indefinite article *ein* is used, the strong masculine ending *-er* appears on the adjective. When the definite article *der*is used, in contrast, the weak ending *-e* appears on the adjective. Thus since the choice of determiner impacts the morphological form of the adjective, there is a morphological dependency pointing from the determiner to the adjective, whereby this morphological dependency is entirely independent of the syntactic dependencies. Consider further the following French sentences:



The masculine subject *le chien* in (a) demands the masculine form of the predicative adjective *blanc*, whereas the feminine subject *la maison* demands the feminine form of this adjective. A morphological dependency that is entirely independent of the syntactic dependencies therefore points again across the syntactic hierarchy.

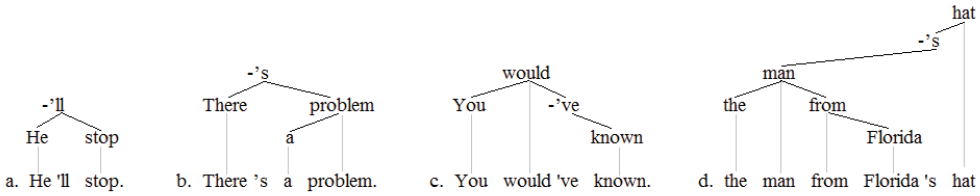
Morphological dependencies play an important role in typological studies. Languages are classified as mostly head-marking (*Sam work-s*) or mostly dependent-marking (*these houses*), whereby most if not all languages contain at least some minor measure of both head and dependent marking.

Prosodic dependencies

Prosodic dependencies are acknowledged in order to accommodate the behavior of clitics. A clitic is a syntactically autonomous element that is prosodically dependent on a host. A cliticis therefore integrated into the prosody of its host, meaning that it forms a single word with its host.

Prosodic dependencies exist entirely in the linear dimension (horizontal dimension), whereas standard syntactic dependencies exist in the hierarchical dimension (vertical dimension). Classic examples of clitics in English are reduced auxiliaries (e.g. *-ll*, *-s*, *-ve*) and the possessive marker *-s*. The

prosodic dependencies in the following examples are indicated with the hyphen and the lack of a vertical projection line:



The hyphens and lack of projection lines indicate prosodic dependencies. A hyphen that appears on the left of the clitic indicates that the clitic is prosodically dependent on the word immediately to its left (*He'll*, *There's*), whereas a hyphen that appears on the right side of the clitic (not shown here) indicates that the clitic is prosodically dependent on the word that appears immediately to its right. A given clitic is often prosodically dependent on its syntactic dependent (*He'll*, *There's*) or on its head (*would've*). At other times, it can depend prosodically on a word that is neither its head nor its immediate dependent (*Florida's*).

Syntactic dependencies

Syntactic dependencies are the focus of most work in DG, as stated above. How the presence and the direction of syntactic dependencies are determined is of course often open to debate. In this regard, it must be acknowledged that the validity of syntactic dependencies in the trees throughout this article is being taken for granted. However, these hierarchies are such that many DGs can largely support them, although there will certainly be points of disagreement. The basic question about how syntactic dependencies are discerned has proven difficult to answer definitively. One should acknowledge in this area,

however, that the basic task of identifying and discerning the presence and direction of the syntactic dependencies of DGs is no easier or harder than determining the constituent groupings of phrase structure grammars. A variety of heuristics are employed to this end, basic tests for constituents being useful tools; the syntactic dependencies assumed in the trees in this article are grouping words together in a manner that most closely matches the results of standard permutation, substitution, and ellipsis tests for constituents. Etymological considerations also provide helpful clues about the direction of dependencies.

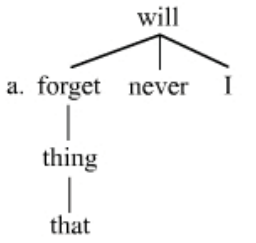
A promising principle upon which to base the existence of syntactic dependencies is distribution. When one is striving to identify the root of a given phrase, the word that is most responsible for determining the distribution of that phrase as a whole is its root.

Linear order and discontinuities

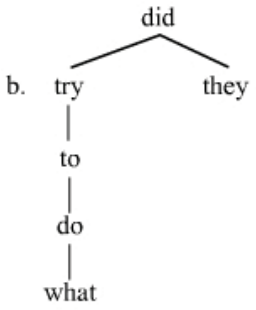
Traditionally, DGs have had a different approach to linear order (word order) than phrase structure grammars. Dependency structures are minimal compared to their phrase structure counterparts, and these minimal structures allow one to focus intently on the two ordering dimensions. Separating the vertical dimension (hierarchical order) from the horizontal dimension (linear order) is easily accomplished.

This aspect of dependency structures has allowed DGs, starting with Tesnière (1959), to focus on hierarchical order in a manner that is hardly possible for phrase structure grammars. For Tesnière, linear order was secondary to

hierarchical order insofar as hierarchical order preceded linear order in the mind of a speaker. The stemmas (trees) that Tesnière produced reflected this view; they abstracted away from linear order to focus almost entirely on hierarchical order. Many DGs that followed Tesnière adopted this practice, that is, they produced tree structures that reflect hierarchical order alone, e.g.



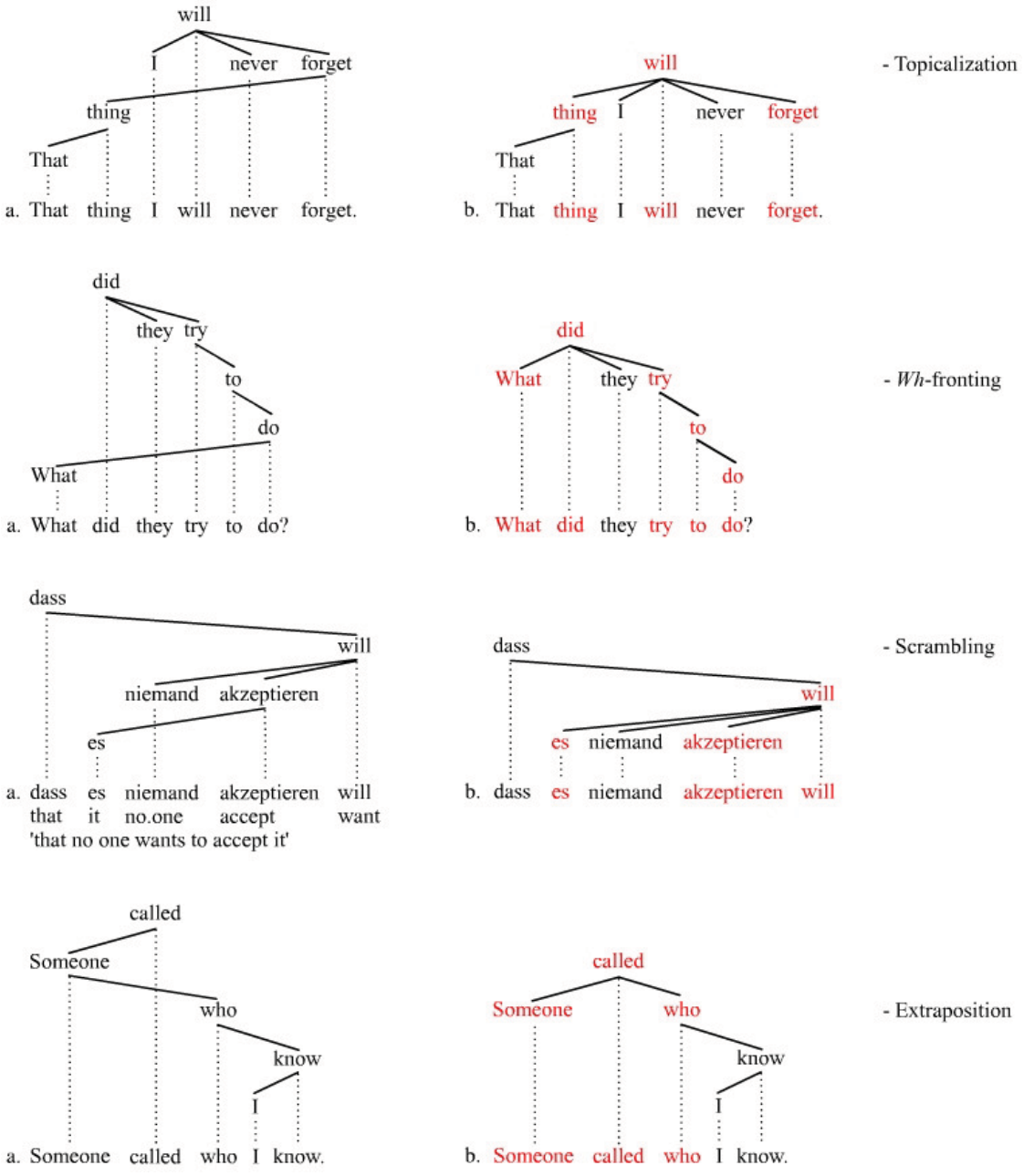
- Unordered tree of the sentence *That thing I will never forget.*



- Unordered tree of the sentence *What did they try to do?.*

The traditional focus on hierarchical order generated the impression that DGs have little to say about linear order, and it has contributed to the view that DGs are particularly well-suited to examine languages with free word order. A negative result of this focus on hierarchical order, however, is that there is a dearth of DG explorations of particular word order phenomena, such as of standard discontinuities. Comprehensive dependency grammar accounts of topicalization, *wh*-fronting, scrambling, and extraposition are mostly absent from many established DG frameworks. This situation can be contrasted with phrase structure grammars, which have devoted tremendous effort to exploring these phenomena.

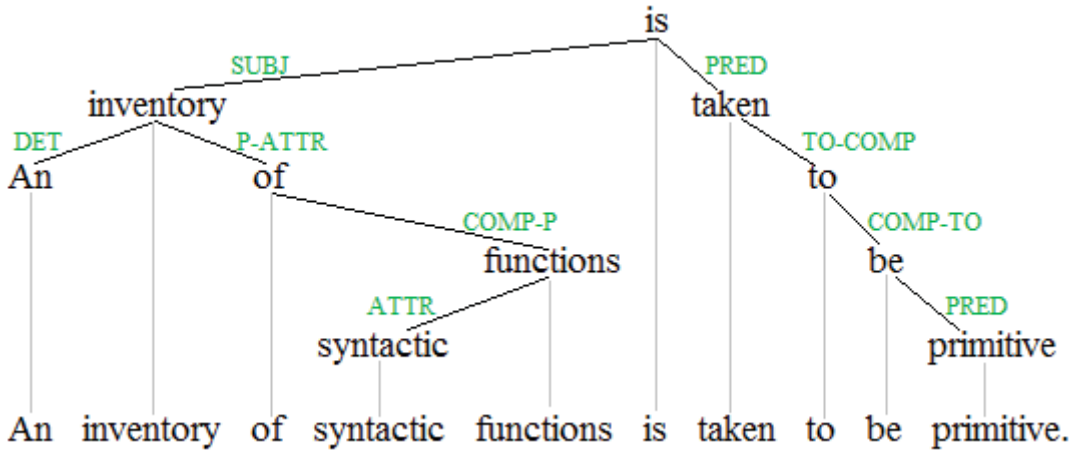
The nature of the dependency relation does not, however, prevent one from focusing on linear order. Dependency structures are as capable of exploring word order phenomena as phrase structures. The following trees illustrate this point; they represent one way of exploring discontinuities using dependency structures. The trees suggest the manner in which common discontinuities can be addressed. An example from German is used to illustrate a scrambling discontinuity:



The a-trees on the left show projectivity violations (= crossing lines), and the b-trees on the right demonstrate one means of addressing these violations. The displaced constituent takes on a word as its head that is not its governor. The words in red mark the catena (=chain) of words that extends from the root of the displaced constituent to the governor of that constituent. Discontinuities are then explored in terms of these catenae. The limitations on topicalization, *wh*-fronting, scrambling, and extraposition can be explored and identified by examining the nature of the catenae involved.

Syntactic functions

Traditionally, DGs have treated the syntactic functions (= grammatical functions, grammatical relations) as primitive. They posit an inventory of functions (e.g. subject, object, oblique, determiner, attribute, predicative, etc.). These functions can appear as labels on the dependencies in the tree structures, e.g.



The syntactic functions in this tree are shown in green: ATTR (attribute), COMP-P (complement of preposition), COMP-TO

(complement of to), DET (determiner), P-ATTR (prepositional attribute), PRED (predicative), SUBJ (subject), TO-COMP (to complement). The functions chosen and abbreviations used in the tree here are merely representative of the general stance of DGs toward the syntactic functions. The actual inventory of functions and designations employed vary from DG to DG.

As a primitive of the theory,the status of these functions is very different from that in some phrase structure grammars. Traditionally, phrase structure grammars derive the syntactic functions from the constellation.

For instance, the object is identified as the NP appearing inside finite VP, and the subject as the NP appearing outside of finite VP. Since DGs reject the existence of a finite VP constituent, they were never presented with the option to view the syntactic functions in this manner. The issue is a question of what comes first: traditionally, DGs take the syntactic functions to be primitive and they then derive the constellation from these functions, whereas phrase structure grammars traditionally take the constellation to be primitive and they then derive the syntactic functions from the constellation.

This question about what comes first (the functions or the constellation) is not an inflexible matter. The stances of both grammar types (dependency and phrase structure) are not narrowly limited to the traditional views. Dependency and phrase structure are both fully compatible with both approaches to the syntactic functions. Indeed, monostratalsystems, that are solely based on dependency or phrase structure, will likely reject the notion that the functions are derived from the constellation or that the

constellation is derived from the functions. They will take both to be primitive, which means neither can be derived from the other.

Categorial grammar

Categorial grammar is an approach in which constituents combine as function and argument, according to combinatory possibilities specified in their syntactic categories. For example, where other approaches might posit a rule that combines a noun phrase (NP) and a verb phrase (VP), CG would posit a syntactic category *NP* and another *NP**S*, read as "a category that searches to the left (indicated by \) for an NP (the element on the left) and outputs a sentence (the element on the right)."

So the syntactic category for an intransitive verb is a complex formula representing the fact that the verb acts as a function word requiring an NP as an input and produces a sentence level structure as an output. This complex category is notated as (NP\S) instead of V.

The category of transitive verbis defined as an element that requires two NPs (its subject and its direct object) to form a sentence.

This is notated as (NP/(NP\S)) which means "a category that searches to the right (indicated by /) for an NP (the object), and generates a function (equivalent to the VP) which is (NP\S), which in turn represents a function that searches to the left for an NP and produces a sentence."

Tree-adjoining grammar is a categorial grammar that adds in partial tree structures to the categories.

Stochastic/probabilistic grammars/network theories

Theoretical approaches to syntax that are based upon probability theory are known as stochastic grammars. One common implementation of such an approach makes use of a neural network or connectionism.

Functional grammars

Functionalist models of grammar study the form–function interaction by performing a structural and a functional analysis.

- Functional discourse grammar (Dik)
- Prague linguistic circle
- Role and reference grammar (RRG)
- Systemic functional grammar

Generative syntax

Generative syntax is the study of syntax within the overarching framework of generative grammar. Generative theories of syntax typically propose analyses of grammatical patterns using formal tools such as phrase structure grammars augmented with additional operations such as syntactic movement. Their goal in analyzing a particular language is to specify rules which generate all and only the expressions which are well-formed in that language. In doing so, they seek to identify innate domain-specific principles of linguistic

cognition, in line with the wider goals of the generative enterprise. Generative syntax is among the approaches that adopt the principle of the autonomy of syntax, assuming that meaning and communicative intent is determined by the syntax rather than the other way around.

Generative syntax was proposed in the late 1950s by Noam Chomsky, building on earlier work by Zellig Harris,Louis Hjelmslev, among others. Since then, numerous theories have been proposed under its umbrella:

- Transformational grammar (TG) (Original theory of generative syntax laid out by Chomsky in *Syntactic Structures* in 1957)
- Government and binding theory (GB) (revised theory in the tradition of TG developed mainly by Chomsky in the 1970s and 1980s)
- Minimalist program (MP) (a reworking of the theory out of the GB framework published by Chomsky in 1995)

Other theories that find their origin in the generative paradigm are:

- Arc pair grammar
- Generalized phrase structure grammar (GPSG)
- Generative semantics
- Head-driven phrase structure grammar (HPSG)
- Lexical functional grammar (LFG)
- Nanosyntax
- Relational grammar (RG)
- Harmonic grammar (HG)

Cognitive and usage-based grammars

The Cognitive Linguistics framework stems from generative grammar, but adheres to evolutionary rather than Chomskyan linguistics. Cognitive models often recognise the generative assumption that the object belongs to the verb phrase. Cognitive frameworks include:

- Cognitive grammar
- Construction grammar (CxG)
- Emergent grammar

Chapter 5

Semantics Deals with the Meaning of Words and Sentences

Semantics (from Ancient Greek: σημαντικός*sēmantikós*, "significant") is the study of meaning, reference, or truth. The term can be used to refer to subfields of several distinct disciplines, including philosophy, linguistics and computer science.

Linguistics

In linguistics, semantics is the subfield that studies meaning. Semantics can address meaning at the levels of words, phrases, sentences, or larger units of discourse. Two of the fundamental issues in the field of semantics are that of compositional semantics (which pertains on how smaller parts, like words, combine and interact to form the meaning of larger expressions such as sentences) and lexical semantics (the nature of the meaning of words). Other prominent issues are those of context and its role on interpretation, opaque contexts, ambiguity, vagueness, entailment and presuppositions.

Several disciplines and approaches have contributed to the often contentious field of semantics. One of the crucial questions which unites different approaches to linguistic semantics is that of the relationship between form and meaning, and some major contributions to the study of

semantics have derived from studies in the 1980-90s in related subjects of the syntax–semantics interface and pragmatics.

The semantic level of language interacts with other modules or levels (like syntax) in which language is traditionally divided. In linguistics, it is typical to talk in terms of "interfaces" regarding such interactions between modules or levels. For semantics, the most crucial interfaces are considered those with semantics (the syntax–semantics interface), pragmatics and phonology (regarding prosody and intonation).

Disciplines and paradigms in linguistic semantics

Formal semantics

Formal semantics seeks to identify domain-specific mental operations which speakers perform when they compute a sentence's meaning on the basis of its syntactic structure. Theories of formal semantics are typically floated on top of theories of syntax such as generative syntax or Combinatory categorial grammar and provide a model theory based on mathematical tools such as typed lambda calculi. The field's central ideas are rooted in early twentieth century philosophical logic as well as later ideas about linguistic syntax. It emerged as its own subfield in the 1970s after the pioneering work of Richard Montague and Barbara Partee and continues to be an active area of research.

Conceptual semantics

Conceptual semantics is a framework for semantic analysis developed mainly by Ray Jackendoff in 1976. Its aim is to

provide a characterization of the conceptual elements by which a person understands words and sentences, and thus to provide *an explanatory semantic representation* (title of a Jackendoff 1976 paper). *Explanatory* in this sense refers to the ability of a given linguistic theory to describe how a component of language is acquired by a child (as proposed by Noam Chomsky; see Levels of adequacy).

Recently, conceptual semantics in particular, and lexical semantics in general, have taken on increasing importance in linguistics and psycholinguistics. Many contemporary theories of syntax (how sentences are constructed from individual words) rely on elements that are idiosyncratic to words themselves. As a result, a sound theory accounting for the properties of the meanings of words is required.

Meaning and decomposition

Jackendoff has claimed that the goal of conceptual semantics is to investigate:

> - "...how linguistic utterances are related to human cognition, where cognition is a human capacity that is to a considerable degree independent of language, interacting with the perceptual and action systems as well as language."
>
> — *(Jackendoff*

Conceptual semantics distinguishes a single, universal meaning to a word. Instead of having a lexical semantic meaning in addition to the conceptual representation of the

actual referent, here the two are combined into what Jackendoff calls "lexical concepts" (Murphy 2010:59). Conceptual semantics is considered to be not just a linguistic theory, but a theory on human cognition. Like many semantic theories, Jackendoff claims that a decompositional method is necessary to explore conceptualization. Just as one of the ways a physical scientist tries to understand matter is by breaking it down into progressively smaller parts, so a scientific study of conceptualization proceeds by breaking down, or decomposing, meanings into smaller parts. However, this decomposition cannot go on forever, for at some point, meanings can no longer be broken down.

This is the level of conceptual structure, the level of mental representations which encode the human understanding of the world, containing the primitive conceptual elements out of which meanings are built, plus their rules of combination. Conceptual semantics does not work with a mental dictionary, in the classical sense. There are no definitions attached to concepts and reference, only the idea of the concept or reference itself. Just as generative syntax posits a finite set of syntactic categories and rules for combining them, so, too, does Conceptual Semantics posit 'a finite set of mental primitives and a finite set of principles of mental combination' governing their interaction (Jackendoff 1990: 9). Jackendoff refers to this set of primitives and the rules governing them as the 'grammar of sentential concepts' (Jackendoff 1990: 9).

His starting point is a close analysis of the meanings of lexemes dedicated to bringing out parallelisms and contrasts which reveal the nature of the conceptual structures underlying them. Jackendoff considers the lexicon to be made

of three parts: phonological, syntactic, and conceptual. These three aspects of a concept give a "full picture of a word" (Murphy 2010:60). What his method shows, he says, is that the psychological organization on which meaning rests 'lies a very short distance below the surface of everyday lexical items – and that progress can be made in exploring it' (1991: 44). Jackendoff claims that a decompositional method is necessary to explore conceptual structure, in which the concepts underlying word meaning are broken down into their smallest elements: conceptual primitives envisaged as the semantic equivalents of phonological features.

Conceptual Semantics posits 'a finite set of mental primitives and a finite set of principles of mental combination' governing their interaction.

The conceptual structure of a lexical item is an element with zero or more open argument slots, which are filled by the syntactic complements of the lexical item.

Semantic structures

Conceptual semantics breaks lexical concepts up into ontological categories: events, states, places, amounts, things, and property, to name a few. These ontological categories are calledsemantic primes, or semantic primitives. Jackendoff poses that any concept in the human brain can be expressed using these semantic primes.

Conceptual semantics is compositional, in that the meanings of phrases, clauses, and sentences can be determined from the lexical concepts that make them up. (Murphy 2010:66)

Problems

Jackendoff's system has been criticised for its highly abstract primitives, which linguists such as Wierzbicka (2007a, 2007b) and Goddard (1998, 2001) have called "obscure". The main reason for this is because one requires special training to understand them, and they often must be translated into plain English to be communicated. Another criticism often raised against conceptual semantics is that it is arbitrary. In its current state, there are no clear procedures for determining when a primitive is justified. Another criticism Wierzbicka and Goddard have raised is that the theory was formulated around and applied only to English, though it claims to be universal.

Jackendoff responds to these criticisms by saying:

> - In fact, an isolated primitive can never be justified: a primitive makes sense only in the context of the overall system of primitives in which it is embedded. With this proviso, however, I think a particular choice of primitives should be justified on the grounds of its capacity for expressing generalizations and explaining the distribution of the data. That is, a proposed system of primitives is subject to the usual scientific standards of evaluation.
> - — *(Jackendoff 1990)*

Cognitive semantics

Cognitive semantics approaches meaning from the perspective of cognitive linguistics. In this framework, language is

explained via general human cognitive abilities rather than a domain-specific language module. The techniques native to cognitive semantics are typically used in lexical studies such as those put forth by Leonard Talmy, George Lakoff, Dirk Geeraerts, and Bruce Wayne Hawkins. Some cognitive semantic frameworks, such as that developed by Talmy, take into account syntactic structures as well. Semantics, through modern researchers can be linked to the Wernicke's area of the brain and can be measured using the event-related potential (ERP). ERP is the rapid electrical response recorded with small disc electrodes which are placed on a person's scalp.

Lexical semantics

A linguistic theory that investigates word meaning. This theory understands that the meaning of a word is fully reflected by its context. Here, the meaning of a word is constituted by its contextual relations. Therefore, a distinction between degrees of participation as well as modes of participation are made. In order to accomplish this distinction any part of a sentence that bears a meaning and combines with the meanings of other constituents is labeled as a semantic constituent. Semantic constituents that cannot be broken down into more elementary constituents are labeled minimal semantic constituents.

Cross-cultural semantics

Various fields or disciplines have long been contributing to cross-cultural semantics. Are words like *love*, *truth*, and *hate* universals? Is even the word *sense* – so central to semantics – a universal, or a concept entrenched in a long-standing but culture-specific tradition? These are the kind of crucial

questions that are discussed in cross-cultural semantics. Translation theory, ethnolinguistics, linguistic anthropology and cultural linguistics specialize in the field of comparing, contrasting, and translating words, terms and meanings from one language to another (see Herder, W. von Humboldt, Boas, Sapir, and Whorf). But philosophy, sociology, and anthropology have long established traditions in contrasting the different nuances of the terms and concepts we use. And online encyclopaedias such as the Stanford encyclopedia of philosophy, Stanford Encyclopedia of Philosophy, and more and more Wikipedia itself have greatly facilitated the possibilities of comparing the background and usages of key cultural terms. In recent years the question of whether key terms are translatable or untranslatable has increasingly come to the fore of global discussions, especially since the publication of Barbara Cassin's *Dictionary of Untranslatables: A Philosophical Lexicon*, in 2014.

Computational semantics

Computational semantics is focused on the processing of linguistic meaning. In order to do this concrete algorithms and architectures are described. Within this framework the algorithms and architectures are also analyzed in terms of decidability, time/space complexity, data structures that they require and communication protocols.

Philosophy

Many of the formal approaches to semantics in mathematical logic and computer science originated in early twentieth

century philosophy of language and philosophical logic. Initially, the most influential semantic theory stemmed from GottlobFrege and Bertrand Russell. Frege and Russell are seen as the originators of a tradition in analytic philosophy to explain meaning compositionally via syntax and mathematical functionality. Ludwig Wittgenstein, a former student of Russell, is also seen as one of the seminal figures in the analytic tradition. All three of these early philosophers of language were concerned with how sentences expressed information in the form of propositions. They also dealt with the truth values or truth conditions a given sentence has in virtue of the proposition it expresses.

In present day philosophy, the term "semantics" is often used to refer to linguistic formal semantics, which bridges both linguistics and philosophy. There is also an active tradition of metasemantics, which studies the foundations of natural language semantics.

Computer science

In computer science, the term *semantics* refers to the meaning of language constructs, as opposed to their form (syntax). According to Euzenat, semantics "provides the rules for interpreting the syntax which do not provide the meaning directly but constrains the possible interpretations of what is declared."

Programming languages

The semantics of programming languages and other languages is an important issue and area of study in computer science.

Like the syntax of a language, its semantics can be defined exactly. For instance, the following statements use different syntaxes, but cause the same instructions to be executed, namely, perform an arithmetical addition of 'y' to 'x' and store the result in a variable called 'x':

Various ways have been developed to describe the semantics of programming languages formally, building on mathematical logic:

- Operational semantics: The meaning of a construct is specified by the computation it induces when it is executed on a machine. In particular, it is of interest *how* the effect of a computation is produced.
- Denotational semantics: Meanings are modelled by mathematical objects that represent the effect of executing the constructs. Thus*only* the effect is of interest, not how it is obtained.
- Axiomatic semantics: Specific properties of the effect of executing the constructs are expressed as *assertions*. Thus there may be aspects of the executions that are ignored.

Semantic models

The Semantic Web refers to the extension of the World Wide Web via embedding added semantic metadata, using semantic data modeling techniques such as Resource Description Framework (RDF) and Web Ontology Language (OWL). On the Semantic Web, terms such as *semantic network* and *semantic data model*are used to describe particular types of data model

characterized by the use of directed graphs in which the vertices denote concepts or entities in the world and their properties, and the arcs denote relationships between them. These can formally be described as description logic concepts and roles, which correspond to OWL classes and properties.

Operational semantics

Operational semantics is a category of formal programming languagesemantics in which certain desired properties of a program, such as correctness, safety or security, are verified by constructing proofs from logical statements about its execution and procedures, rather than by attaching mathematical meanings to its terms (denotational semantics).

Operational semantics are classified in two categories: **structural operational semantics** (or **small-step semantics**) formally describe how the *individual steps* of a computation take place in a computer-based system; by opposition**natural semantics** (or **big-step semantics**) describe how the *overall results* of the executions are obtained.

Other approaches to providing a formal semantics of programming languages include axiomatic semantics and denotational semantics.

The operational semantics for a programming language describes how a valid program is interpreted as sequences of computational steps. These sequences then *are* the meaning of the program. In the context of functional programming, the final step in a terminating sequence returns the value of the program. (In general there can be many return values for a

single program, because the program could be nondeterministic, and even for a deterministic program there can be many computation sequences since the semantics may not specify exactly what sequence of operations arrives at that value.)

Perhaps the first formal incarnation of operational semantics was the use of the lambda calculus to define the semantics of Lisp. Abstract machines in the tradition of the SECD machine are also closely related.

History

The concept of operational semantics was used for the first time in defining the semantics of Algol 68.

The following statement is a quote from the revised ALGOL 68 report:

The meaning of a program in the strict language is explained in terms of a hypothetical computer which performs the set of actions that constitute the elaboration of that program. (Algol68, Section 2)

The first use of the term "operational semantics" in its present meaning is attributed to Dana Scott (Plotkin04).

What follows is a quote from Scott's seminal paper on formal semantics, in which he mentions the "operational" aspects of semantics.

It is all very well to aim for a more 'abstract' and a 'cleaner' approach to semantics, but if the plan is to be any good, the operational aspects cannot be completely ignored. (Scott70)

Approaches

Gordon Plotkin introduced the structural operational semantics, Matthias Felleisen and Robert Hieb the reduction semantics, and Gilles Kahn the natural semantics.

Small-step semantics

Structural operational semantics

- **Structural operational semantics** (SOS, also called **structured operational semantics** or **small-step semantics**) was introduced by Gordon Plotkin in (Plotkin81) as a logical means to define operational semantics. The basic idea behind SOS is to define the behavior of a program in terms of the behavior of its parts, thus providing a structural, i.e., syntax-oriented and inductive, view on operational semantics. An SOS specification defines the behavior of a program in terms of a (set of) transition relation(s). SOS specifications take the form of a set of inference rules that define the valid transitions of a composite piece of syntax in terms of the transitions of its components.

Big-step semantics

Natural semantics

Big-step structural operational semantics is also known under the names natural semantics, relational semantics and evaluation semantics. Big-step operational semantics was introduced under the name *natural semantics* by Gilles Kahn when presenting Mini-ML, a pure dialect of ML.

One can view big-step definitions as definitions of functions, or more generally of relations, interpreting each language construct in an appropriate domain. Its intuitiveness makes it a popular choice for semantics specification in programming languages, but it has some drawbacks that make it inconvenient or impossible to use in many situations, such as languages with control-intensive features or concurrency.

A big-step semantics describes in a divide-and-conquer manner how final evaluation results of language constructs can be obtained by combining the evaluation results of their syntactic counterparts (subexpressions, substatements, etc.).

Comparison

There are a number of distinctions between small-step and big-step semantics that influence whether one or the other forms a more suitable basis for specifying the semantics of a programming language.

Big-step semantics have the advantage of often being simpler (needing fewer inference rules) and often directly correspond to

an efficient implementation of an interpreter for the language (hence Kahn calling them "natural".) Both can lead to simpler proofs, for example when proving the preservation of correctness under some program transformation.

The main disadvantage of big-step semantics is that non-terminating (diverging) computations do not have an inference tree, making it impossible to state and prove properties about such computations.

Small-step semantics give more control over the details and order of evaluation. In the case of instrumented operational semantics, this allows the operational semantics to track and the semanticist to state and prove more accurate theorems about the run-time behaviour of the language. These properties make small-step semantics more convenient when proving type soundness of a type system against an operational semantics.

Denotational semantics

In computer science, denotational semantics (initially known as mathematical semantics or Scott–Strachey semantics) is an approach of formalizing the meanings of programming languages by constructing mathematical objects (called *denotations*) that describe the meanings of expressions from the languages. Other approaches providing formal semantics of programming languages include axiomatic semantics and operational semantics.

Broadly speaking, denotational semantics is concerned with finding mathematical objects called domains that represent what programs do. For example, programs (or program

phrases) might be represented by partial functions or by games between the environment and the system.

An important tenet of denotational semantics is that *semantics should be compositional*: the denotation of a program phrase should be built out of the denotations of its subphrases.

Historical development

Denotational semantics originated in the work of Christopher Strachey and Dana Scott published in the early 1970s. As originally developed by Strachey and Scott, denotational semantics provided the meaning of a computer program as a function that mapped input into output. To give meanings to recursively defined programs, Scott proposed working with continuous functions between domains, specifically complete partial orders. As described below, work has continued in investigating appropriate denotational semantics for aspects of programming languages such as sequentiality, concurrency, non-determinism and local state.

Denotational semantics has been developed for modern programming languages that use capabilities like concurrency and exceptions, e.g., Concurrent ML, CSP, and Haskell. The semantics of these languages is compositional in that the meaning of a phrase depends on the meanings of its subphrases. For example, the meaning of the applicative expressionf(E1,E2) is defined in terms of semantics of its subphrases f, E1 and E2. In a modern programming language, E1 and E2 can be evaluated concurrently and the execution of one of them might affect the other by interacting through shared objects causing their meanings to be defined in terms of

each other. Also, E1 or E2 might throw an exception which could terminate the execution of the other one. The sections below describe special cases of the semantics of these modern programming languages.

Meanings of recursive programs

Denotational semantics is ascribed to a program phrase as a function from an environment (holding the current values of its free variables) to its denotation. For example, the phrase `n*m` produces a denotation when provided with an environment that has binding for its two free variables: `n` and `m`. If in the environment`n` has the value 3 and `m` has the value 5, then the denotation is 15.

A function can be represented as a set of ordered pairs of argument and corresponding result values. For example, the set {(0,1), (4,3)} denotes a function with result 1 for argument 0, result 3 for the argument 4, and undefined otherwise.

The problem to be solved is to provide meanings for recursive programs that are defined in terms of themselves such as the definition of the factorial function as

```
intfactorial(intn){if(n==0)thenreturn1;elsereturnn*factorial(n-1);}
```

A solution is to build up the meanings by approximation. The factorial function is a total function from $\mathbb{N}$ to $\mathbb{N}$ (defined everywhere in its domain), but we model it as a partial function. At the beginning, we start with the empty function (an empty set). Next, we add the ordered pair (0,1) to the function to result in another partial function that better approximates the factorial function. Afterwards, we add yet

another ordered pair (1,1) to create an even better approximation. It is instructive to think of this chain of iteration for a "partial factorial function" F as F, F, F, ... where F indicates F applied n times.

- $F(\{\})$ is the totally undefined partial function, represented as the set {};
- $F(\{\})$ is the partial function represented as the set {(0,1)}: it is defined at 0, to be 1, and undefined elsewhere;
- $F(\{\})$ is the partial function represented as the set {(0,1), (1,1), (2,2), (3,6), (4,24)}: it is defined for arguments 0,1,2,3,4.

Denotational semantics of non-deterministic programs

The concept of power domainshas been developed to give a denotational semantics to non-deterministic sequential programs. Writing P for a power-domain constructor, the domain $P(D)$ is the domain of non-deterministic computations of type denoted by D.

There are difficulties with fairness and unboundedness in domain-theoretic models of non-determinism.

Denotational semantics of concurrency

Many researchers have argued that the domain-theoretic models given above do not suffice for the more general case of concurrent computation. For this reason various new models have been introduced. In the early 1980s, people began using the style of denotational semantics to give semantics for

concurrent languages. Examples include Will Clinger's work with the actor model; Glynn Winskel's work with event structures and petri nets; and the work by Francez, Hoare, Lehmann, and de Roever (1979) on trace semantics for CSP. All these lines of inquiry remain under investigation (see e.g. the various denotational models for CSP).

Recently, Winskel and others have proposed the category of profunctors as a domain theory for concurrency.

Denotational semantics of state

State (such as a heap) and simple imperative featurescan be straightforwardly modeled in the denotational semantics described above. All the textbooks below have the details. The key idea is to consider a command as a partial function on some domain of states. The meaning of "`x:=3`" is then the function that takes a state to the state with `3` assigned to `x`. The sequencing operator "`;`" is denoted by composition of functions. Fixed-point constructions are then used to give a semantics to looping constructs, such as "`while`".

Things become more difficult in modelling programs with local variables. One approach is to no longer work with domains, but instead to interpret types as functors from some category of worlds to a category of domains. Programs are then denoted by natural continuous functions between these functors.

Denotations of data types

Many programming languages allow users to define recursive data types. For example, the type of lists of numbers can be specified by

```
datatypelist=Consofnat*list | Empty
```

This section deals only with functional data structures that cannot change. Conventional imperative programming languages would typically allow the elements of such a recursive list to be changed.

For another example: the type of denotations of the untyped lambda calculus is

```
datatypeD=Dof (D→D)
```

The problem of *solving domain equations* is concerned with finding domains that model these kinds of datatypes. One approach, roughly speaking, is to consider the collection of all domains as a domain itself, and then solve the recursive definition there. The textbooks below give more details.

Polymorphic data types are data types that are defined with a parameter. For example, the type of α`list`s is defined by

```
datatypeαlist=Consofα*αlist|Empty
```

Lists of natural numbers, then, are of type `nat list`, while lists of strings are of `string list`.

Some researchers have developed domain theoretic models of polymorphism. Other researchers have also modeled parametric polymorphism within constructive set theories. Details are found in the textbooks listed below.

A recent research area has involved denotational semantics for object and class based programming languages.

Denotational semantics for programs of restricted complexity

Following the development of programming languages based on linear logic, denotational semantics have been given to languages for linear usage (see e.g. proof nets, coherence spaces) and also polynomial time complexity.

Denotational semantics of sequentiality

The problem of full abstraction for the sequential programming language PCF was, for a long time, a big open question in denotational semantics. The difficulty with PCF is that it is a very sequential language. For example, there is no way to define the parallel-or function in PCF. It is for this reason that the approach using domains, as introduced above, yields a denotational semantics that is not fully abstract.

This open question was mostly resolved in the 1990s with the development of game semanticsand also with techniques involving logical relations. For more details, see the page on PCF.

Denotational semantics as source-to-source translation

It is often useful to translate one programming language into another. For example, a concurrent programming language might be translated into a process calculus; a high-level programming language might be translated into byte-code. (Indeed, conventional denotational semantics can be seen as the interpretation of programming languages into the internal language of the category of domains.)

In this context, notions from denotational semantics, such as full abstraction, help to satisfy security concerns.

Abstraction

It is often considered important to connect denotational semantics with operational semantics. This is especially important when the denotational semantics is rather mathematical and abstract, and the operational semantics is more concrete or closer to the computational intuitions. The following properties of a denotational semantics are often of interest.

- **Syntax independence**: The denotations of programs should not involve the syntax of the source language.
- **Adequacy (or soundness)**: All observably distinct programs have distinct denotations;
- **Full abstraction**: All observationally equivalent programs have equal denotations.

For semantics in the traditional style, adequacy and full abstraction may be understood roughly as the requirement that "operational equivalence coincides with denotational equality". For denotational semantics in more intensional models, such as the actor model and process calculi, there are different notions of equivalence within each model, and so the concepts of adequacy and of full abstraction are a matter of debate, and harder to pin down. Also the mathematical structure of operational semantics and denotational semantics can become very close.

Additional desirable properties we may wish to hold between operational and denotational semantics are:

- **Constructivism**: Constructivism is concerned with whether domain elements can be shown to exist by constructive methods.
- **Independence of denotational and operational semantics**: The denotational semantics should be formalized using mathematical structures that are independent of the operational semantics of a programming language; However, the underlying concepts can be closely related. See the section on Compositionality below.
- **Full completeness** or **definability**: Every morphism of the semantic model should be the denotation of a program.

Connections to other areas of computer science

Some work in denotational semantics has interpreted types as domains in the sense of domain theory, which can be seen as a branch of model theory, leading to connections with type theory and category theory.

Within computer science, there are connections with abstract interpretation, program verification, and model checking.

Psychology

Semantic memory

In psychology, *semantic memory* is memory for meaning – in other words, the aspect of memory that preserves only the *gist*, the general significance, of remembered experience – while episodic memory is memory for the ephemeral details – the individual features, or the unique particulars of experience. The term 'episodic memory' was introduced by Tulving and Schacter in the context of 'declarative memory' which involved simple association of factual or objective information concerning its object. Word meaning is measured by the company they keep, i.e. the relationships among words themselves in a semantic network.

The memories may be transferred intergenerationally or isolated in one generation due to a cultural disruption. Different generations may have different experiences at similar points in their own time-lines. This may then create a vertically heterogeneous semantic net for certain words in an otherwise homogeneous culture. In a network created by people analyzing their understanding of the word (such as Wordnet) the links and decomposition structures of the network are few in number and kind, and include *part of*, *kind of*, and similar links. In automated ontologies the links are computed vectors without explicit meaning. Various automated technologies are being developed to compute the meaning of words: latent semantic indexing and support vector machines as well as natural language processing, artificial neural networks and predicate calculus techniques.

Ideasthesia

Ideasthesia is a psychological phenomenon in which activation of concepts evokes sensory experiences. For example, in synesthesia, activation of a concept of a letter (e.g., that of the letter *A*) evokes sensory-like experiences (e.g., of red color).

Psychosemantics

In the 1960s, psychosemantic studies became popular after Charles E. Osgood's massive cross-cultural studies using his semantic differential (SD) method that used thousands of nouns and adjective bipolar scales. A specific form of the SD, Projective Semantics method uses only most common and neutral nouns that correspond to the 7 groups (factors) of adjective-scales most consistently found in cross-cultural studies (Evaluation, Potency, Activity as found by Osgood, and Reality, Organization, Complexity, Limitation as found in other studies). In this method, seven groups of bipolar adjective scales corresponded to seven types of nouns so the method was thought to have the object-scale symmetry (OSS) between the scales and nouns for evaluation using these scales. For example, the nouns corresponding to the listed 7 factors would be: Beauty, Power, Motion, Life, Work, Chaos, Law. Beauty was expected to be assessed unequivocally as "very good" on adjectives of Evaluation-related scales, Life as "very real" on Reality-related scales, etc. However, deviations in this symmetric and very basic matrix might show underlying biases of two types: scales-related bias and objects-related bias. This OSS design meant to increase the sensitivity of the SD method to any semantic biases in responses of people within the same culture and educational background.

Prototype theory

Another set of concepts related to fuzziness in semantics is based on prototypes. The work of Eleanor Rosch in the 1970s led to a view that natural categories are not characterizable in terms of necessary and sufficient conditions, but are graded (fuzzy at their boundaries) and inconsistent as to the status of their constituent members. One may compare it with Jung's archetype, though the concept of archetype sticks to static concept. Some post-structuralists are against the fixed or static meaning of the words. Derrida, following Nietzsche, talked about slippages in fixed meanings.

Systems of categories are not objectively *out there* in the world but are rooted in people's experience. These categories evolve as learned concepts of the world – meaning is not an objective truth, but a subjective construct, learned from experience, and language arises out of the "grounding of our conceptual systems in shared embodiment and bodily experience". A corollary of this is that the conceptual categories (i.e. the lexicon) will not be identical for different cultures, or indeed, for every individual in the same culture. This leads to another debate (see the Sapir–Whorf hypothesis or Eskimo words for snow).

Semantic technology

The ultimate goal of **semantic technology** is to help machines understand data. To enable the encoding of semantics with the data, well-known technologies are RDF (Resource Description Framework) and OWL (Web Ontology Language). These

technologies formally represent the meaning involved in information. For example, ontology can describe concepts, relationships between things, and categories of things. These embedded semantics with the data offer significant advantages such as reasoning over data and dealing with heterogeneous data sources.

Overview

In software, semantic technology encodes meanings separately from data and content files, and separately from application code. This enables machines as well as people to understand, share and reason with them at execution time. With semantic technologies, adding, changing and implementing new relationships or interconnecting programs in a different way can be just as simple as changing the external model that these programs share.

With traditional information technology, on the other hand, meanings and relationships must be predefined and "hard wired" into data formats and the application program code at design time. This means that when something changes, previously unexchanged information needs to be exchanged, or two programs need to interoperate in a new way, the humans must get involved.

Off-line, the parties must define and communicate between them the knowledge needed to make the change, and then recode the data structures and program logic to accommodate it, and then apply these changes to the database and the application. Then, and only then, can they implement the changes.

Semantic technologies are "meaning-centered". They involve but are not limited to the following areas of application:

- encoding/decoding of semantic representation,
- knowledge graphs of entities and their interrelationships,
- auto-recognition of topics and concepts,
- information and meaning extraction,
- semantic data integration, and
- taxonomies/classification.

Given a question, semantic technologies can directly search topics, concepts, associations that span a vast number of sources.

Semantic technologies provide an abstraction layer above existing IT technologies that enables bridging and interconnection of data, content, and processes. Second, from the portal perspective, semantic technologies can be thought of as a new level of depth that provides far more intelligent, capable, relevant, and responsive interaction than with information technologies alone.