# Impact of Linguistics

Gabriel Poole

# IMPACT OF LINGUISTICS

# IMPACT OF LINGUISTICS

Gabriel Poole

BIBLIOTEX
Digital Library

Impact of Linguistics
by Gabriel Poole



www.bibliotex.com



Ebook ISBN: 9781984665379


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

# Contents

Chapter 1

# Applied Linguistics

**Applied linguistics** is an interdisciplinary field which identifies, investigates, and offers solutions to language-related real-life problems. Some of the academic fields related to applied linguistics are education, psychology, communication research, anthropology, and sociology.

# Domain

Applied linguistics is an interdisciplinary field. Major branches of applied linguistics include bilingualism and multilingualism, conversation analysis, contrastive linguistics, sign linguistics, language assessment, literacies, discourse analysis, language pedagogy, second language acquisition, language planning and policy, interlinguistics, stylistics, language teacher education, pragmatics, forensic linguistics and translation.

# Journals

Major journals of the field include *Annual Review of Applied Linguistics*, *Applied Linguistics*, Studies in Second Language Acquisition, *Applied Psycholinguistics*, *International Review of Applied Linguistics in Language Teaching*, *International Journal of Applied Linguistics*, *Applied Linguistics Review*, *European Journal of Applied Linguistics*, *Language Learning*, *Language*

*and Education*, System, *TESOL Quarterly*, and *Linguistics and Education*.

See also List of applied linguistics journals

# History

The tradition of applied linguistics established itself in part as a response to the narrowing of focus in linguistics with the advent in the late 1950s of generative linguistics, and has always maintained a socially-accountable role, demonstrated by its central interest in language problems.

Although the field of applied linguistics started from Europe and the United States, the field rapidly flourished in the international context.

Applied linguistics first concerned itself with principles and practices on the basis of linguistics. In the early days, applied linguistics was thought as “linguistics-applied” at least from the outside of the field. In the 1960s, however, applied linguistics was expanded to include language assessment, language policy, and second language acquisition. As early as the 1970s, applied linguistics became a problem-driven field rather than theoretical linguistics, including the solution of language-related problems in the real world. By the 1990s, applied linguistics had broadened including critical studies and multilingualism. Research in applied linguistics was shifted to "the theoretical and empirical investigation of real world problems in which language is a central issue."

In the United States, applied linguistics also began narrowly as the application of insights from structural linguistics—first to the teaching of English in schools and subsequently to second and foreign language teaching. The *linguistics applied* approach to language teaching was promulgated most strenuously by Leonard Bloomfield, who developed the foundation for the Army Specialized Training Program, and by Charles C. Fries, who established the English Language Institute (ELI) at the University of Michigan in 1941. In 1946, Applied linguistics became a recognized field of studies in the aforementioned university. In 1948, the Research Club at Michigan established *Language Learning: A Journal of Applied Linguistics*, the first journal to bear the term *applied linguistics.* In the late 1960s, applied linguistics began to establish its own identity as an interdisciplinary field of linguistics concerned with real-world language issues. The new identity was solidified by the creation of the American Association for Applied Linguistics in 1977.

# Associations

The International Association of Applied Linguisticswas founded in France in 1964, where it is better known as Association Internationale de Linguistique Appliquée, or AILA. AILA has affiliates in more than thirty countries, some of which are listed below.

## Australia

Australian applied linguistics took as its target the applied linguistics of mother tongue teaching and teaching English to

immigrants. The Australia tradition shows a strong influence of continental Europe and of the US, rather than of Britain. Applied Linguistics Association of Australia (ALAA) was established at a national congress of applied linguists held in August 1976. ALAA holds a joint annual conference in collaboration with the Association for Applied Linguistics in New Zealand (ALANZ).

## Canada

The Canadian Association of Applied Linguistics / L'Associationcanadienne de linguistiqueappliquée (CAAL/ACLA), is an officially bilingual (English and French) scholarly association with approximately 200 members. They produce the *Canadian Journal of Applied Linguistics* and hold an annual conference.

## Ireland

The Irish Association for Applied Linguistics/CumannnaTeangeolaíochtaFeidhmí (IRAAL) was founded in 1975. They produce the journal *Teanga*, the Irish word for 'language'.

## Japan

In 1982, the Japan Association of Applied Linguistics (JAAL) was established in the Japan Association of College English Teachers (JACET) in order to engage in activities on a more international scale. In 1984, JAAL became an affiliate of the International Association of Applied Linguistics (AILA).[1]

## New Zealand

The Applied Linguistics Association of New Zealand (ALANZ) produces the journal *New Zealand Studies in Applied Linguistics* and has been collaborating with the Applied Linguistics Association of Australia in a combined annual conference since 2010, with the Association for Language Testing and Assessment of Australia and New Zealand (ALTAANZ) later joining the now three-way conference collaboration.

## South Africa

The Southern African Applied Linguistics Association (SAALA) was founded in 1980. There are currently four publications associated with SAALA including the *Southern African Linguistics and Applied Language Studies Journal* (SAJALS).

## United Kingdom

The British Association for Applied Linguistics (BAAL) was established in 1967. Its mission is "the advancement of education by fostering and promoting, by any lawful charitable means, the study of language use, language acquisition and language teaching and the fostering of interdisciplinary collaboration in this study [...]". BAAL hosts an annual conference, as well as many additional smaller conferences and events organised by its Special Interest Groups (SIGs).

## United States

The American Association for Applied Linguistics (AAAL) was founded in 1977. AAAL holds an annual conference, usually in March or April, in the United States or Canada.

# 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's Aufbau, 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's competition 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 buildingmodel 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.

# Anthropological linguistics

**Anthropological linguistics** is the subfield of linguistics and anthropology, which deals with the place of language in its wider social and cultural context, and its role in making and maintaining cultural practices and societal structures. While many linguists believe that a true field of anthropological linguistics is nonexistent, preferring the term linguistic anthropology to cover this subfield, many others regard the two as interchangeable.

# History

Although researchers studied the two fields together at various points in the nineteenth century, the intersection of anthropology and linguistics significantly grew in prominence during the early twentieth century. As American scholarship became increasingly interested in the diversity of Native American societies in the New World, anthropologists and linguists worked in conjunction to analyze Native American languages and to study how language related to the origins, distribution, and characteristics of these indigenous populations.

This interdisciplinary approach distinguished American anthropology from its European counterpart; while European anthropology largely focused on ethnography, American anthropology began to integrate linguistics and other disciplines. Anthropological linguistics initially focused largely

on unwritten language, but now examines languages both with and without written traditions.

Early anthropological linguists primarily focused on three major areas: linguistic description, classification, and methodology.

- **Linguistic Description:** Scholars such as Franz Boas, Edward Sapir, Leonard Bloomfield, and Mary Haas drafted descriptions of linguistic structure and the linguistic characteristics of different languages. They conducted research as fieldwork, using recordings of texts from native speakers and performing analysis to categorize the texts by linguistic form and genre.
- **Classification:** Classification involved outlining the genetic relationships among languages. Linguistic classifications allowed anthropological linguists to organize large amounts of information about specific populations. By classifying language, scholars could systematize and order data from their ethnographic work.
- **Methodology:** By analytically breaking down language, anthropological linguistics could use the constituent parts to derive social and cultural information. It also made pattern-identification possible, with Boas and Sapir using these procedures to show that linguistic patterning was unrealized among speakers of a given language.

# Overview

Anthropological linguistics is one of many disciplines which studies the role of languages in the social lives of individuals and within communities. To do this, experts have had to understand not only the logic behind linguistic systems – such as their grammars – but also record the activities in which those systems are used. In the 1960s and 1970s, sociolinguistics and anthropological linguistics were often viewed as one single field of study, but they have since become more separate as more academic distance has been put between them. Though there are many similarities and a definite sharing of topics – such as gender and language – they are two related but separate entities. Anthropological linguistics came about in the United States as a subfield of anthropology, when anthropologists were beginning to study the indigenous cultures, and the indigenous languages could no longer be ignored, and quickly morphed into the subfield of linguistics that it is known as today.

Anthropological linguistics has had a major impact in the studies of such areas as visual perception (especially colour) and bioregional democracy, both of which are concerned with distinctions that are made in languages about perceptions of the surroundings.

Conventional linguistic anthropology also has implications for sociology and self-organization of peoples. Study of the Penan people, for instance, reveals that their language employs six different and distinct words whose best English translation is "we". Anthropological linguistics studies these distinctions,

and relates them to types of societies and to actual bodily adaptation to the senses, much as it studies distinctions made in languages regarding the colours of the rainbow: seeing the tendency to increase the diversity of terms, as evidence that there are distinctions that bodies in this environment *must* make, leading to situated knowledge and perhaps a situated ethics, whose final evidence is the differentiated set of terms used to denote "we".

The two branches of anthropological linguistics are nomenclatural/classificational and ethnographic/sociolinguistics.

Indexicality refers to language forms that is tied to meaning through association of specific and general, as opposed to direct naming. For example, an anthropological linguist may utilize indexicality to analyze what an individual's use of language reveals about his or her social class. Indexicality is inherent in form-function relationships.

# Distinction from Other Subfields

Although the terms anthropological linguistics and linguistic anthropology are often viewed as being synonymous, specialists often make a distinction between them. While anthropological linguistics is considered a subfield of linguistics, linguistic anthropology is generally considered to be a subfield of anthropology. Anthropological linguistics also uses more distinctly linguistic methodology, and studies languages as "linguistic phenomena." Ultimately, anthropological linguistics focuses on the cultural and social

meaning of language, with more of an emphasis on linguistic structure. Conversely, linguistic anthropology uses more anthropological methods (such as participant observation and fieldwork) to analyze language through a cultural framework and determine the rules of its social use.

While anthropological linguistics uses language to determine cultural understandings, sociolinguistics views language itself as a social institution. Anthropological linguistics is largely interpretative, striving to determine the significance behind the use of language through its forms, registers, and styles. Sociolinguistics instead examines how language relates to various social groups and identities like race, gender, class, and age.

# Structures

## Phonology

A common variation of linguistics that focuses on the sounds within speech of any given language. It outlines why phonetic features identify words.

Phonology puts a large focus on the systematic structure of the sounds being observed.

## Morphology

Morphology in linguistics commonly looks at the structure of words within a language to develop a better understanding for

the word formbeing used. It is the branch of linguistics that deals with words, their internal structure, and their formation. Morphology looks broadly at the connection of word forms within a specific language in relation to the culture or environment it is rooted within.

## Methodology

There are two major trends in the theoretical and methodological study of attitudes in the social sciences - mentalist and behaviorist. The mentalist trend treats attitude as a mediating concept while the behaviorist trend operationally defines it as a probability concept, though in research practice both derive their attitude measures from response variation. While there are many different views concerning the structure and components of attitudes, there is, however, an overwhelming agreement that attitudes are learned, lasting, and positively related to behavior. Methodology in attitude studies includes direct and indirect measures of all kinds, but language attitude studies have tended to make more use of questionnaires than of other methods. The matched guise technique - a sociolinguistic experimental technique used to determine the true feelings of an individual or community towards a specific language, dialect, or accent - has been extensively used for studies relating to the social significance of languages and language varieties. A special adaptation of this technique, called mirror image, appears promising for measuring consensual evaluations of language switching at the situational level. Situational based self-report instruments such as those used by Greenfield and Fishman also promise to be very effective instruments for studies pertaining to normative views

concerning the situational use of languages and language varieties. The commitment measure has been found to be particularly suited for collecting data on behavioral tendencies. Data obtained through interviewing may be difficult to process and score – and may provide bias from those being interviewed – but the research interview can be particularly effective for attitude assessment, especially when used to complement the observational method. Data collected through the observational method can be formally processed like data obtained through more formalized instruments if attempts are made to record the data in more public forms instead of only through the approach most characteristic for this kind of data have used so far.

Many linguists believe that comparisons of linguistic and social behaviorhave been blocked by the fact that linguistic and anthropological studies are rarely based on comparable sets of data. While an anthropologist's description refers to specific communities, linguistic analysis refers to a single language or dialect, and the behaviors formed through verbal signs and structural similarities. The process of linguistic analysis is oriented towards the discovery of unitary, structurally similar wholes. The effect of these procedures is the selection of one single variety out of the many varieties that characterize everyday speech and behavior. Englishis often thought of as one single language, as though people forget the many dialects and accents that come with it. English spoken in the United States of America will not be the same English spoken in Australia, or in the countries of Africa. Even American English spoken in New York will not be exactly the same as American English spoken in Alabama.

# Code-switching

While code-switching, a situation in which a speaker alternates between two or more languages, or language varieties, in the context of a single conversation, is not the only form of linguistic variability to carry a social, or referential meaning, it does provide a particularly clear approach to understanding the relationship between social processes and linguistic forms, because both the social and the linguistic boundaries in question tend to be most evident than in other monolingual settings. In anthropological linguistics, code-switching has been approached as a structurally unified phenomenon whose significance comes from a universal pattern of relationships between form, function, and context. Many linguists are approaching code-switching as a form of verbal strategy, which represents the ways in which the linguistic resources available to individuals may vary according to the nature of their social boundaries within their communities. While the emphasis is on language use in social interaction as the preferred focus for examining exactly how those processes work, it is clear that future research must take into account the situation of that interaction within the specific community, or across communities. The study of code-switching will increasingly be able to contribute to an understanding of the nature of speech communities.

# Related fields

Anthropological linguistics is concerned with

- Descriptive (or synchronic) linguistics: Describing dialects (forms of a language used by a specific speech community). This study includes phonology, morphology, syntax, semantics, and grammar.
- Historical (or diachronic) linguistics: Describing changes in dialects and languages over time. This study includes the study of linguistic divergence and language families, comparative linguistics, etymology, and philology.
- Ethnolinguistics: Analyzing the relationship between culture, thought, and language.
- Sociolinguistics: Analyzing the social functions of language and the social, political, and economic relationships among and between members of speech communities.

## Chapter 2

# Discourse Analysis

**Discourse analysis** (**DA**), or **discourse studies**, is an approach to the analysis of written, vocal, or sign language use, or any significant semiotic event.

The objects of discourse analysis (discourse, writing, conversation, communicative event) are variously defined in terms of coherent sequences of sentences, propositions, speech, or turns-at-talk. Contrary to much of traditional linguistics, discourse analysts not only study language use 'beyond the sentence boundary' but also prefer to analyze 'naturally occurring' language use, not invented examples. Text linguistics is a closely related field. The essential difference between discourse analysis and text linguistics is that discourse analysis aims at revealing socio-psychological characteristics of a person/persons rather than text structure.

Discourse analysis has been taken up in a variety of disciplines in the humanities and social sciences, including linguistics, education, sociology, anthropology, social work, cognitive psychology, social psychology, area studies, cultural studies, international relations, human geography, environmental science, communication studies, biblical studies, public relations, argumentation studies, and translation studies, each of which is subject to its own assumptions, dimensions of analysis, and methodologies.

# History

## Early use of the term

The ancient Greeks (among others) had much to say on discourse; however, there is ongoing discussion about whether Austria-born Leo Spitzer's *Stilstudien* (*Style Studies*) of 1928 the earliest example of *discourse analysis* (DA). Michel Foucault translated it into French. However, the term first came into general use following the publication of a series of papers by Zellig Harris from 1952 reporting on work from which he developed transformational grammar in the late 1930s. Formal equivalence relations among the sentences of a coherent discourse are made explicit by using sentence transformations to put the text in a canonical form. Words and sentences with equivalent information then appear in the same column of an array.

This work progressed over the next four decades (see references) into a science of sublanguage analysis (Kittredge &Lehrberger 1982), culminating in a demonstration of the informational structures in texts of a sublanguage of science, that of Immunology, (Harris et al. 1989) and a fully articulated theory of linguistic informational content (Harris 1991). During this time, however, most linguists ignored such developments in favor of a succession of elaborate theories of sentence-level syntax and semantics.

In January 1953, a linguist working for the American Bible Society, James A. Lauriault/Loriot, needed to find answers to some fundamental errors in translating Quechua, in the Cuzco

area of Peru. Following Harris's 1952 publications, he worked over the meaning and placement of each word in a collection of Quechua legends with a native speaker of Quechua and was able to formulate discourse rules that transcended the simple sentence structure. He then applied the process to Shipibo, another language of Eastern Peru. He taught the theory at the Summer Institute of Linguistics in Norman, Oklahoma, in the summers of 1956 and 1957 and entered the University of Pennsylvania to study with Harris in the interim year. He tried to publish a paper *Shipibo Paragraph Structure*, but it was delayed until 1970 (Loriot&Hollenbach 1970). In the meantime, Kenneth Lee Pike, a professor at University of Michigan, Ann Arbor, taught the theory, and one of his students, Robert E. Longacre developed it in his writings. Harris's methodology disclosing the correlation of form with meaning was developed into a system for the computer-aided analysis of natural language by a team led by Naomi Sager at NYU, which has been applied to a number of sublanguage domains, most notably to medical informatics. The software for the Medical Language Processor is publicly available on SourceForge.

## In the humanities

In the late 1960s and 1970s, and without reference to this prior work, a variety of other approaches to a new cross-discipline of DA began to develop in most of the humanities and social sciences concurrently with, and related to, other disciplines. These include semiotics, psycholinguistics, sociolinguistics, and pragmatics. Many of these approaches, especially those influenced by the social sciences, favor a more dynamic study of oral talk-in-interaction. An example is

"conversational analysis", which was influenced by the Sociologist Harold Garfinkel, the founder of Ethnomethodology.

## Foucault

In Europe, Michel Foucault became one of the key theorists of the subject, especially of discourse, and wrote The Archaeology of Knowledge. In this context, the term 'discourse' no longer refers to formal linguistic aspects, but to institutionalized patterns of knowledge that become manifest in disciplinary structures and operate by the connection of knowledge and power. Since the 1970s, Foucault's works have had an increasing impact especially on discourse analysis in the field of social sciences. Thus, in modern European social sciences, one can find a wide range of different approaches working with Foucault's definition of discourse and his theoretical concepts. Apart from the original context in France, there is, since 2005, a broad discussion on socio-scientific discourse analysis in Germany. Here, for example, the sociologistReiner Keller developed his widely recognized 'Sociology of Knowledge Approach to Discourse (SKAD)'. Following the sociology of knowledge by Peter L. Berger and Thomas Luckmann, Keller argues that our sense of reality in everyday life and thus the meaning of every object, action and event is the product of a permanent, routinized interaction. In this context, SKAD has been developed as a scientific perspective that is able to understand the processes of 'The Social Construction of Reality' on all levels of social life by combining the prementioned Michel Foucault's theories of discourse and power while also introducing the theory of knowledge by Berger/Luckmann. Whereas the latter primarily focus on the constitution and stabilization of knowledge on the level of

interaction, Foucault's perspective concentrates on institutional contexts of the production and integration of knowledge, where the subject mainly appears to be determined by knowledge and power. Therefore, the 'Sociology of Knowledge Approach to Discourse' can also be seen as an approach to deal with the vividly discussed micro–macro problem in sociology.

# Perspectives

The following are some of the specific theoretical perspectives and analytical approaches used in linguistic discourse analysis:

- Applied linguistics, an interdisciplinary perspective on linguistic analysis
- Cognitive neuroscience of discourse comprehension
- Cognitive psychology, studying the production and comprehension of discourse.
- Conversation analysis
- Critical discourse analysis
- Discursive psychology
- Emergent grammar
- Ethnography of communication
- Functional grammar
- Interactional sociolinguistics
- Mediated Stylistics
- Pragmatics
- Response based therapy (counselling)
- Rhetoric
- Stylistics (linguistics)

- Sublanguage analysis
- Tagmemics
- Text linguistics
- Variation analysis

Although these approaches emphasize different aspects of language use, they all view language as social interaction and are concerned with the social contexts in which discourse is embedded.

Often a distinction is made between 'local' structures of discourse (such as relations among sentences, propositions, and turns) and 'global' structures, such as overall topics and the schematic organization of discourses and conversations. For instance, many types of discourse begin with some kind of global 'summary', in titles, headlines, leads, abstracts, and so on.

A problem for the discourse analyst is to decide when a particular feature is relevant to the specification required. A question many linguists ask is: "Are there general principles which will determine the relevance or nature of the specification?"

# Topics of interest

Topics of discourse analysis include:

- The various levels or dimensions of discourse, such as sounds (intonation, etc.), gestures, syntax, the lexicon, style, rhetoric, meanings, speech acts,

moves, strategies, turns, and other aspects of interaction

- Genres of discourse (various types of discourse in politics, the media, education, science, business, etc.)
- The relations between discourse and the emergence of syntactic structure
- The relations between text (discourse) and context
- The relations between discourse and power
- The relations between discourse and interaction
- The relations between discourse and cognition and memory

# Prominent academics

- Jan Blommaert
- Teun van Dijk
- Michel Foucault
- Heidi E. Hamilton
- Barbara Johnstone
- SinfreeMakoni
- Jonathan Potter
- Deborah Schiffrin
- Deborah Tannen
- Margaret Wetherell
- Ruth Wodak

# Political discourse

Political Discourse: The text and talk of professional politicians or political institutions, such as presidents and prime ministers and other members of government, parliament or political parties, both at the local, national and international levels, includes both the speaker and the audience.

Political discourse analysis is a field of discourse analysis which focuses on discourse in political forums (such as debates, speeches, and hearings) as the phenomenon of interest. Policy analysis requires discourse analysis to be effective from the post-positivist perspective.

Political discourse is the formal exchange of reasoned views as to which of several alternative courses of action should be taken to solve a societal problem.

# Corporate discourse

Corporate discourse can be broadly defined as the language used by corporations. It encompasses a set of messages that a corporation sends out to the world (the general public, the customers and other corporations) and the messages it uses to communicate within its own structures (the employees and other stakeholders).

# Computational linguistics

**Computational linguistics** is an interdisciplinary field concerned with the computational modelling of natural language, as well as the study of appropriate computational approaches to linguistic questions. In general, computational linguistics draws upon linguistics, computer science, artificial intelligence, mathematics, logic, philosophy, cognitive science, cognitive psychology, psycholinguistics, anthropology and neuroscience, among others.

# Sub-fields and related areas

Traditionally, computational linguistics emerged as an area of artificial intelligence performed by computer scientists who had specialized in the application of computers to the processing of a natural language. With the formation of the Association for Computational Linguistics (ACL) and the establishment of independent conference series, the field consolidated during the 1970s and 1980s.

The Association for Computational Linguistics defines computational linguistics as:

...the scientific study of language from a computational perspective. Computational linguists are interested in providing computational models of various kinds of linguistic phenomena.

The term "computational linguistics" is nowadays (2020) taken to be a near-synonym of natural language processing (NLP) and (human) language technology. These terms put a stronger emphasis on aspects of practical applications rather than theoretical inquiry and since the 2000s. In practice, they have largely replaced the term "computational linguistics" in the NLP/ACL community, although they specifically refer to the sub-field of applied computational linguistics, only.

Computational linguistics has both theoretical and applied components. Theoretical computational linguistics focuses on issues in theoretical linguistics and cognitive science. Applied computational linguistics focuses on the practical outcome of modeling human language use.

**Theoretical computational linguistics** includes the development of formal theories of grammar (parsing) and semantics, often grounded in formal logics and symbolic (knowledge-based) approaches. Areas of research that are studied by theoretical computational linguistics include:

- Computational complexity of natural language, largely modeled on automata theory, with the application of context-sensitive grammar and linearly boundedTuring machines.
- Computational semantics comprises defining suitable logics for linguistic meaning representation, automatically constructing them and reasoning with them

**Applied computational linguistics** is dominated by machine learning, traditionally using statistical methods, since the mid-2010s by neural networks: Socher et al. (2012) was an early

Deep Learning tutorial at the ACL 2012, and met with both interest and (at the time) scepticism by most participants. Until then, neural learning was basically rejected because of its lack of statistical interpretability. Until 2015, deep learning had evolved into the major framework of NLP. As for the tasks addressed by applied computational linguistics, see Natural language processing article. This includes classical problems such as the design of POS-taggers (part-of-speech taggers), parsers for natural languages, or tasks such as machine translation (MT), the sub-division of computational linguistics dealing with having computers translate between languages. As one of the earliest and most difficult applications of computational linguistics, MT draws on many subfields and both theoretical and applied aspects. Traditionally, automatic language translation has been considered a notoriously hard branch of computational linguistics.

Aside from dichothomy between theoretical and applied computational linguistics, other divisions of computational into major areas according to different criteria exist, including:

- **medium** of the language being processed, whether spoken or textual: speech recognition and speech synthesis deal with how spoken language can be understood or created using computers.
- **task** being performed, e.g., whether analyzing language (recognition) or synthesizing language (generation): Parsing and generation are sub-divisions of computational linguistics dealing respectively with taking language apart and putting it together.

Traditionally, applications of computers to address research problems in other branches of linguistics have been described as tasks within computational linguistics. Among other aspects, this includes

- Computer-aided corpus linguistics, which has been used since the 1970s as a way to make detailed advances in the field of discourse analysis
- Simulation and study of language evolution in historical linguistics/glottochronology.

# Origins

Computational linguistics is often grouped within the field of artificial intelligence but was present before the development of artificial intelligence. Computational linguistics originated with efforts in the United States in the 1950s to use computers to automatically translate texts from foreign languages, particularly Russian scientific journals, into English. Since computers can make arithmetic (systematic) calculations much faster and more accurately than humans, it was thought to be only a short matter of time before they could also begin to process language. Computational and quantitative methods are also used historically in the attempted reconstruction of earlier forms of modern languages and sub-grouping modern languages into language families. Earlier methods, such as lexicostatistics and glottochronology, have been proven to be premature and inaccurate. However, recent interdisciplinary studies that borrow concepts from biological studies, especially gene mapping, have proved to produce more sophisticated analytical tools and more reliable results.

When machine translation (also known as mechanical translation) failed to yield accurate translations right away, automated processing of human languages was recognized as far more complex than had originally been assumed. Computational linguistics was born as the name of the new field of study devoted to developing algorithms and software for intelligently processing language data. The term "computational linguistics" itself was first coined by David Hays, a founding member of both the Association for Computational Linguistics (ACL) and the International Committee on Computational Linguistics (ICCL).

To translate one language into another, it was observed that one had to understand the grammar of both languages, including both morphology (the grammar of word forms) and syntax (the grammar of sentence structure). To understand syntax, one had to also understand the semantics and the lexicon (or 'vocabulary'), and even something of the pragmatics of language use. Thus, what started as an effort to translate between languages evolved into an entire discipline devoted to understanding how to represent and process natural languages using computers.

Nowadays research within the scope of computational linguistics is done at computational linguistics departments, computational linguistics laboratories, computer science departments, and linguistics departments. Some research in the field of computational linguistics aims to create working speech or text processing systems while others aim to create a system allowing human-machine interaction. Programs meant for human-machine communication are calledconversational agents.

# Approaches

Just as computational linguistics can be performed by experts in a variety of fields and through a wide assortment of departments, so too can the research fields broach a diverse range of topics. The following sections discuss some of the literature available across the entire field broken into four main area of discourse: developmental linguistics, structural linguistics, linguistic production, and linguistic comprehension.

## Developmental approaches

Language is a cognitive skill that develops throughout the life of an individual. This developmental process has been examined using several techniques, and a computational approach is one of them. Human language development does provide some constraints which make it harder to apply a computational method to understanding it. For instance, during language acquisition, human children are largely only exposed to positive evidence. This means that during the linguistic development of an individual, the only evidence for what is a correct form is provided, and no evidence for what is not correct. This is insufficient information for a simple hypothesis testing procedure for information as complex as language, and so provides certain boundaries for a computational approach to modeling language development and acquisition in an individual.

Attempts have been made to model the developmental process of language acquisition in children from a computational angle,

leading to both statistical grammars and connectionist models. Work in this realm has also been proposed as a method to explain the evolution of language through history. Using models,it has been shown that languages can be learned with a combination of simple input presented incrementally as the child develops better memory and longer attention span. This was simultaneously posed as a reason for the long developmental period of human children. Both conclusions were drawn because of the strength of the artificial neural network which the project created.

The ability of infants to develop language has also been modeled using robots in order to test linguistic theories. Enabled to learn as children might, a model was created based on an affordance model in which mappings between actions, perceptions, and effects were created and linked to spoken words. Crucially, these robots were able to acquire functioning word-to-meaning mappings without needing grammatical structure, vastly simplifying the learning process and shedding light on information which furthers the current understanding of linguistic development. It is important to note that this information could only have been empirically tested using a computational approach.

As our understanding of the linguistic development of an individual within a lifetime is continually improved using neural networks and learning robotic systems, it is also important to keep in mind that languages themselves change and develop through time. Computational approaches to understanding this phenomenon have unearthed very interesting information. Using the Price equation and Pólya urn dynamics, researchers have created a system which not only

predicts future linguistic evolution but also gives insight into the evolutionary history of modern-day languages. This modeling effort achieved, through computational linguistics, what would otherwise have been impossible.

It is clear that the understanding of linguistic development in humans as well as throughout evolutionary time has been fantastically improved because of advances in computational linguistics. The ability to model and modify systems at will affords science an ethical method of testing hypotheses that would otherwise be intractable.

## Structural approaches

To create better computational models of language, an understanding of language's structure is crucial. To this end, the English languagehas been meticulously studied using computational approaches to better understand how the language works on a structural level. One of the most important pieces of being able to study linguistic structure is the availability of large linguistic corpora or samples. This grants computational linguists the raw data necessary to run their models and gain a better understanding of the underlying structures present in the vast amount of data which is contained in any single language. One of the most cited English linguistic corpora is the Penn Treebank. Derived from widely-different sources, such as IBM computer manuals and transcribed telephone conversations, this corpus contains over 4.5 million words of American English. This corpus has been primarily annotated using part-of-speech tagging and syntactic bracketing and has yielded substantial empirical observations related to language structure.

Theoretical approaches to the structure of languages have also been developed. These works allow computational linguistics to have a framework within which to work out hypotheses that will further the understanding of the language in a myriad of ways. One of the original theoretical theses on the internalization of grammar and structure of language proposed two types of models. In these models, rules or patterns learned increase in strength with the frequency of their encounter. The work also created a question for computational linguists to answer: how does an infant learn a specific and non-normal grammar (Chomsky normal form) without learning an overgeneralized version and getting stuck? Theoretical efforts like these set the direction for research to go early in the lifetime of a field of study, and are crucial to the growth of the field.

Structural information about languages allows for the discovery and implementation of similarity recognition between pairs of text utterances. For instance, it has recently been proven that based on the structural information present in patterns of human discourse, conceptual recurrence plots can be used to model and visualize trends in data and create reliable measures of similarity between natural textual utterances. This technique is a strong tool for further probing the structure of human discourse. Without the computational approach to this question, the vastly complex information present in discourse data would have remained inaccessible to scientists.

Information regarding the structural data of a language is available for English as well as other languages, such as Japanese. Using computational methods, Japanese sentence

corpora were analyzed and a pattern of log-normality was found in relation to sentence length. Though the exact cause of this lognormality remains unknown, it is precisely this sort of information which computational linguistics is designed to uncover. This information could lead to further important discoveries regarding the underlying structure of Japanese and could have any number of effects on the understanding of Japanese as a language. Computational linguistics allows for very exciting additions to the scientific knowledge base to happen quickly and with very little room for doubt.

Without a computational approach to the structure of linguistic data, much of the information that is available now would still be hidden under the vastness of data within any single language. Computational linguistics allows scientists to parse huge amounts of data reliably and efficiently, creating the possibility for discoveries unlike any seen in most other approaches.

## Production approaches

- The production of language is equally as complex in the information it provides and the necessary skills which a fluent producer must have. That is to say, comprehension is only half the problem of communication. The other half is how a system produces language, and computational linguistics has made interesting discoveries in this area. In a now famous paper published in 1950Alan Turing proposed the possibility that machines might one day have the ability to "think". As a thought experiment for what might define the concept of

> thought in machines, he proposed an "imitation test" in which a human subject has two text-only conversations, one with a fellow human and another with a machine attempting to respond like a human. Turing proposes that if the subject cannot tell the difference between the human and the machine, it may be concluded that the machine is capable of thought. Today this test is known as the Turing test and it remains an influential idea in the area of artificial intelligence.

One of the earliest and best-known examples of a computer program designed to converse naturally with humans is the ELIZA program developed by Joseph Weizenbaum at MIT in 1966. The program emulated a Rogerianpsychotherapist when responding to written statements and questions posed by a user. It appeared capable of understanding what was said to it and responding intelligently, but in truth, it simply followed a pattern matching routine that relied on only understanding a few keywords in each sentence. Its responses were generated by recombining the unknown parts of the sentence around properly translated versions of the known words. For example, in the phrase "It seems that you hate me" ELIZA understands "you" and "me" which matches the general pattern "you [some words] me", allowing ELIZA to update the words "you" and "me" to "I" and "you" and replying "What makes you think I hate you?". In this example ELIZA has no understanding of the word "hate", but it is not required for a logical response in the context of this type of psychotherapy.

Some projects are still trying to solve the problem which first started computational linguistics off as its field in the first

place. However, methods have become more refined, and consequently, the results generated by computational linguists have become more enlightening. To improve computer translation, several models have been compared, including hidden Markov models, smoothing techniques, and the specific refinements of those to apply them to verb translation. The model which was found to produce the most natural translations of German and French words was a refined alignment model with a first-order dependence and a fertility model. They also provide efficient training algorithms for the models presented, which can give other scientists the ability to improve further on their results. This type of work is specific to computational linguistics and has applications that could vastly improve understanding of how language is produced and comprehended by computers.

Work has also been done in making computers produce language in a more naturalistic manner. Using linguistic input from humans, algorithms have been constructed which are able to modify a system's style of production based on a factor such as linguistic input from a human, or more abstract factors like politeness or any of the five main dimensions of personality. This work takes a computational approach via parameter estimation models to categorize the vast array of linguistic styles we see across individuals and simplify it for a computer to work in the same way, making human–computer interaction much more natural.

## Text-based interactive approach

Many of the earliest and simplest models of human–computer interaction, such as ELIZA for example, involve a text-based input from the user to generate a response from the computer. By this method, words typed by a user trigger the computer to recognize specific patterns and reply accordingly, through a process known as keyword spotting.

## Speech-based interactive approach

Recent technologies have placed more of an emphasis on speech-based interactive systems. These systems, such as Siri of the iOS operating system, operate on a similar pattern-recognizing technique as that of text-based systems, but with the former, the user input is conducted through speech recognition. This branch of linguistics involves the processing of the user's speech as sound waves and the interpreting of the acoustics and language patterns for the computer to recognize the input.

## Comprehension approaches

Much of the focus of modern computational linguistics is on comprehension. With the proliferation of the internet and the abundance of easily accessible written human language, the ability to create a program capable of understanding human language would have many broad and exciting possibilities, including improved search engines, automated customer service, and online education.

Early work in comprehension included applying Bayesian statistics to the task of optical character recognition, as illustrated by Bledsoe and Browing in 1959 in which a large dictionary of possible letters was generated by "learning" from example letters and then the probability that any one of those learned examples matched the new input was combined to make a final decision. Other attempts at applying Bayesian statistics to language analysis included the work of Mosteller and Wallace (1963) in which an analysis of the words used in *The Federalist Papers* was used to attempt to determine their authorship (concluding that Madison most likely authored the majority of the papers).

In 1971Terry Winograd developed an early natural language processing engine capable of interpreting naturally written commands within a simple rule-governed environment. The primary language parsing program in this project was called SHRDLU, which was capable of carrying out a somewhat natural conversation with the user giving it commands, but only within the scope of the toy environment designed for the task. This environment consisted of different shaped and colored blocks, and SHRDLU was capable of interpreting commands such as "Find a block which is taller than the one you are holding and put it into the box." and asking questions such as "I don't understand which pyramid you mean." in response to the user's input. While impressive, this kind of natural language processing has proven much more difficult outside the limited scope of the toy environment. Similarly, a project developed by NASA called LUNARwas designed to provide answers to naturally written questions about the geological analysis of lunar rocks returned by the Apollo

missions. These kinds of problems are referred to as question answering.

Initial attempts at understanding spoken language were based on work done in the 1960s and 1970s in signal modeling where an unknown signal is analyzed to look for patterns and to make predictions based on its history. An initial and somewhat successful approach to applying this kind of signal modeling to language was achieved with the use of hidden Markov models as detailed by Rabiner in 1989. This approach attempts to determine probabilities for the arbitrary number of models that could be being used in generating speech as well as modeling the probabilities for various words generated from each of these possible models. Similar approaches were employed in early speech recognition attempts starting in the late 70s at IBM using word/part-of-speech pair probabilities.

More recently these kinds of statistical approaches have been applied to more difficult tasks such as topic identification using Bayesian parameter estimation to infer topic probabilities in text documents.

# Applications

Applied computational linguistics is largely equivalent with natural language processing. Example applications for end users include speech recognition software, such as Apple's Siri feature, spellcheck tools, speech synthesis programs, which are often used to demonstrate pronunciation or help the disabled, and machine translation programs and websites, such as Google Translate.

Computational linguistics are also helpful in situations involving social media and the Internet, e.g., for providing content filters in chatrooms or on website searches, for grouping and organizing content through social media mining, document retrieval and clustering. For instance, if a person searches "red, large, four-wheeled vehicle," to find pictures of a red truck, the search engine will still find the information desired by matching words such as "four-wheeled" with "car".

Computational approaches are also important to support linguistic research, e.g., in corpus linguistics or historical linguistics. As for the study of change over time, computational methods can contribute to the modeling and identification of language families (see further quantitative comparative linguistics or phylogenetics), as well as the modeling of changes in sound and meaning.

# Legacy

The subject of computational linguistics has had a recurring impact on popular culture:

- The Star Trek franchise features heavily classical NLP applications, most notably machine translation (universal translator), natural language user interfaces and question answering.
- The 1983 film *WarGames* features a young computer hacker who interacts with an artificially intelligent supercomputer.
- A 1997 film, *Conceiving Ada*, focuses on Ada Lovelace, considered one of the first computer

programmers, as well as themes of computational linguistics.

- *Her*, a 2013 film, depicts a man's interactions with the "world's first artificially intelligent operating system."
- The 2014 film *The Imitation Game* follows the life of computer scientist Alan Turing, developer of the Turing Test.
- The 2015 film *Ex Machina* centers around human interaction with artificial intelligence.
- The 2016 film *Arrival*, based on Ted Chiang's Story of Your Life, takes a whole new approach of linguistics to communicate with advanced alien race called heptapods.

Chapter 3

# Neurolinguistics

**Neurolinguistics** is the study of the neural mechanisms in the human brain that control the comprehension, production, and acquisition of language. As an interdisciplinary field, neurolinguistics draws methods and theories from fields such as neuroscience, linguistics, cognitive science, communication disorders and neuropsychology. Researchers are drawn to the field from a variety of backgrounds, bringing along a variety of experimental techniques as well as widely varying theoretical perspectives. Much work in neurolinguistics is informed by models in psycholinguistics and theoretical linguistics, and is focused on investigating how the brain can implement the processes that theoretical and psycholinguistics propose are necessary in producing and comprehending language. Neurolinguists study the physiological mechanisms by which the brain processes information related to language, and evaluate linguistic and psycholinguistic theories, using aphasiology, brain imaging, electrophysiology, and computer modeling.

# History

Neurolinguistics is historically rooted in the development in the 19th century of aphasiology, the study of linguistic deficits (aphasias) occurring as the result of brain damage. Aphasiology attempts to correlate structure to function by analyzing the effect of brain injuries on language processing.

One of the first people to draw a connection between a particular brain area and language processing was Paul Broca, a French surgeon who conducted autopsies on numerous individuals who had speaking deficiencies, and found that most of them had brain damage (or *lesions*) on the left frontal lobe, in an area now known as Broca's area. Phrenologists had made the claim in the early 19th century that different brain regions carried out different functions and that language was mostly controlled by the frontal regions of the brain, but Broca's research was possibly the first to offer empirical evidence for such a relationship, and has been described as "epoch-making" and "pivotal" to the fields of neurolinguistics and cognitive science. Later, Carl Wernicke, after whom Wernicke's areais named, proposed that different areas of the brain were specialized for different linguistic tasks, with Broca's area handling the motor production of speech, and Wernicke's area handling auditory speech comprehension. The work of Broca and Wernicke established the field of aphasiology and the idea that language can be studied through examining physical characteristics of the brain. Early work in aphasiology also benefited from the early twentieth-century work of KorbinianBrodmann, who "mapped" the surface of the brain, dividing it up into numbered areas based on each area's cytoarchitecture (cell structure) and function; these areas, known as Brodmann areas, are still widely used in neuroscience today.

The coining of the term "neurolinguistics" is attributed to Edith Crowell Trager, Henri Hecaen and Alexandr Luria, in the late 1940s and 1950s; Luria's book "Problems in Neurolinguistics" is likely the first book with Neurolinguistics in the title. Harry Whitaker popularized neurolinguistics in the

United States in the 1970s, founding the journal "Brain and Language" in 1974.

Although aphasiology is the historical core of neurolinguistics, in recent years the field has broadened considerably, thanks in part to the emergence of new brain imaging technologies (such as PET and fMRI) and time-sensitive electrophysiological techniques (EEG and MEG), which can highlight patterns of brain activation as people engage in various language tasks; electrophysiological techniques, in particular, emerged as a viable method for the study of language in 1980 with the discovery of the N400, a brain response shown to be sensitive to semantic issues in language comprehension. The N400 was the first language-relevant event-related potential to be identified, and since its discovery EEG and MEG have become increasingly widely used for conducting language research.

# Discipline

## Interaction with other fields

Neurolinguistics is closely related to the field of psycholinguistics, which seeks to elucidate the cognitive mechanisms of language by employing the traditional techniques of experimental psychology; today, psycholinguistic and neurolinguistic theories often inform one another, and there is much collaboration between the two fields.

Much work in neurolinguistics involves testing and evaluating theories put forth by psycholinguists and theoretical linguists. In general, theoretical linguists propose models to explain the

structure of language and how language information is organized, psycholinguists propose models and algorithms to explain how language information is processed in the mind, and neurolinguists analyze brain activity to infer how biological structures (populations and networks of neurons) carry out those psycholinguistic processing algorithms. For example, experiments in sentence processing have used the ELAN, N400, and P600 brain responses to examine how physiological brain responses reflect the different predictions of sentence processing models put forth by psycholinguists, such as Janet Fodor and Lyn Frazier's "serial" model, and Theo Vosse and Gerard Kempen's "unification model". Neurolinguists can also make new predictions about the structure and organization of language based on insights about the physiology of the brain, by "generalizing from the knowledge of neurological structures to language structure".

## Topics considered

Neurolinguistics research investigates several topics, including where language information is processed, how language processing unfolds over time, how brain structures are related to language acquisition and learning, and how neurophysiology can contribute to speech and language pathology.

# Localizations of language processes

Much work in neurolinguistics has, like Broca's and Wernicke's early studies, investigated the locations of specific language "modules" within the brain. Research questions include what course language information follows through the brain as it is

processed, whether or not particular areas specialize in processing particular sorts of information, how different brain regions interact with one another in language processing, and how the locations of brain activation differ when a subject is producing or perceiving a language other than his or her first language.

## Time course of language processes

Another area of neurolinguistics literature involves the use of electrophysiological techniques to analyze the rapid processing of language in time. The temporal ordering of specific patterns of brain activity may reflect discrete computational processes that the brain undergoes during language processing; for example, one neurolinguistic theory of sentence parsing proposes that three brain responses (the ELAN, N400, and P600) are products of three different steps in syntactic and semantic processing.

## Language acquisition

Another topic is the relationship between brain structures and language acquisition. Research in first language acquisition has already established that infants from all linguistic environments go through similar and predictable stages (such as babbling), and some neurolinguistics research attempts to find correlations between stages of language development and stages of brain development, while other research investigates the physical changes (known as neuroplasticity) that the brain undergoes during second language acquisition, when adults

learn a new language. Neuroplasticity is observed when both Second Language acquisition and Language Learning experience are induced, the result of this language exposure concludes that an increase of gray and white matter could be found in children, young adults and the elderly.

Ping Li, Jennifer Legault, Kaitlyn A. Litcofsky, May 2014. Neuroplasticity as a function of second language learning: Anatomical changes in the human brain Cortex: A Journal Devoted to the Study of the Nervous System & Behavior, 410.1016/j.cortex.2014.05.00124996640

## Language pathology

Neurolinguistic techniques are also used to study disorders and breakdowns in language, such as aphasia and dyslexia, and how they relate to physical characteristics of the brain.

# Technology used

Since one of the focuses of this field is the testing of linguistic and psycholinguistic models, the technology used for experiments is highly relevant to the study of neurolinguistics. Modern brain imaging techniques have contributed greatly to a growing understanding of the anatomical organization of linguistic functions. Brain imaging methods used in neurolinguistics may be classified into hemodynamic methods, electrophysiological methods, and methods that stimulate the cortex directly.

## Hemodynamic

Hemodynamic techniques take advantage of the fact that when an area of the brain works at a task, blood is sent to supply that area with oxygen (in what is known as the Blood Oxygen Level-Dependent, or BOLD, response). Such techniques include PET and fMRI. These techniques provide high *spatial resolution*, allowing researchers to pinpoint the location of activity within the brain; *temporal resolution* (or information about the timing of brain activity), on the other hand, is poor, since the BOLD response happens much more slowly than language processing. In addition to demonstrating which parts of the brain may subserve specific language tasks or computations, hemodynamic methods have also been used to demonstrate how the structure of the brain's language architecture and the distribution of language-related activation may change over time, as a function of linguistic exposure.

In addition to PET and fMRI, which show which areas of the brain are activated by certain tasks, researchers also use diffusion tensor imaging (DTI), which shows the neural pathways that connect different brain areas, thus providing insight into how different areas interact. Functional near-infrared spectroscopy (fNIRS) is another hemodynamic method used in language tasks.

## Electrophysiological

Electrophysiological techniques take advantage of the fact that when a group of neurons in the brain fire together, they create an electric dipole or current. The technique of EEG measures

this electric current using sensors on the scalp, while MEG measures the magnetic fields that are generated by these currents. In addition to these non-invasive methods, electrocorticography has also been used to study language processing. These techniques are able to measure brain activity from one millisecond to the next, providing excellent *temporal resolution*, which is important in studying processes that take place as quickly as language comprehension and production. On the other hand, the location of brain activity can be difficult to identify in EEG; consequently, this technique is used primarily to *how* language processes are carried out, rather than *where*. Research using EEG and MEG generally focuses on event-related potentials (ERPs), which are distinct brain responses (generally realized as negative or positive peaks on a graph of neural activity) elicited in response to a particular stimulus. Studies using ERP may focus on each ERP's *latency* (how long after the stimulus the ERP begins or peaks), *amplitude* (how high or low the peak is), or *topography* (where on the scalp the ERP response is picked up by sensors). Some important and common ERP components include the N400 (a negativity occurring at a latency of about 400 milliseconds), the mismatch negativity, the early left anterior negativity (a negativity occurring at an early latency and a front-left topography), the P600, and the lateralized readiness potential.

# Experimental design

## Experimental techniques

Neurolinguists employ a variety of experimental techniques in order to use brain imaging to draw conclusions about how language is represented and processed in the brain. These techniques include the *subtraction* paradigm, *mismatch design*, *violation-based* studies, various forms of *priming*, and *direct stimulation* of the brain.

## Subtraction

Many language studies, particularly in fMRI, use the subtraction paradigm, in which brain activation in a task thought to involve some aspect of language processing is compared against activation in a baseline task thought to involve similar non-linguistic processes but not to involve the linguistic process. For example, activations while participants read words may be compared to baseline activations while participants read strings of random letters (in attempt to isolate activation related to lexical processing—the processing of real words), or activations while participants read syntactically complex sentences may be compared to baseline activations while participants read simpler sentences.

# Mismatch paradigm

The mismatch negativity (MMN) is a rigorously documented ERP component frequently used in neurolinguistic experiments. It is an electrophysiological response that occurs in the brain when a subject hears a "deviant" stimulus in a set of perceptually identical "standards" (as in the sequence *s ssssss d d s sssss d s ssss d*). Since the MMN is elicited only in response to a rare "oddball" stimulus in a set of other stimuli that are perceived to be the same, it has been used to test how speakers perceive sounds and organize stimuli categorically. For example, a landmark study by Colin Phillips and colleagues used the mismatch negativity as evidence that subjects, when presented with a series of speech sounds with acoustic parameters, perceived all the sounds as either /t/ or /d/ in spite of the acoustic variability, suggesting that the human brain has representations of abstract phonemes—in other words, the subjects were "hearing" not the specific acoustic features, but only the abstract phonemes. In addition, the mismatch negativity has been used to study syntactic processing and the recognition of word category.

# Violation-based

Many studies in neurolinguistics take advantage of anomalies or *violations* of syntactic or semantic rules in experimental stimuli, and analyzing the brain responses elicited when a subject encounters these violations. For example, sentences beginning with phrases such as **the garden was on the worked*, which violates an English phrase structure rule, often

elicit a brain response called the early left anterior negativity (ELAN). Violation techniques have been in use since at least 1980, when Kutas and Hillyard first reported ERP evidence that semantic violations elicited an N400 effect. Using similar methods, in 1992, Lee Osterhout first reported the P600 response to syntactic anomalies. Violation designs have also been used for hemodynamic studies (fMRI and PET): Embick and colleagues, for example, used grammatical and spelling violations to investigate the location of syntactic processing in the brain using fMRI. Another common use of violation designs is to combine two kinds of violations in the same sentence and thus make predictions about how different language processes interact with one another; this type of crossing-violation study has been used extensively to investigate how syntactic and semantic processes interact while people read or hear sentences.

# Priming

In psycholinguistics and neurolinguistics, *priming* refers to the phenomenon whereby a subject can recognize a word more quickly if he or she has recently been presented with a word that is similar in meaning or morphological makeup (i.e., composed of similar parts). If a subject is presented with a "prime" word such as *doctor* and then a "target" word such as *nurse*, if the subject has a faster-than-usual response time to *nurse* then the experimenter may assume that word *nurse* in the brain had already been accessed when the word *doctor* was accessed. Priming is used to investigate a wide variety of questions about how words are stored and retrieved in the brain and how structurally complex sentences are processed.

# Stimulation

Transcranial magnetic stimulation (TMS), a new noninvasive technique for studying brain activity, uses powerful magnetic fields that are applied to the brain from outside the head. It is a method of exciting or interrupting brain activity in a specific and controlled location, and thus is able to imitate aphasic symptoms while giving the researcher more control over exactly which parts of the brain will be examined. As such, it is a less invasive alternative to direct cortical stimulation, which can be used for similar types of research but requires that the subject's scalp be removed, and is thus only used on individuals who are already undergoing a major brain operation (such as individuals undergoing surgery for epilepsy). The logic behind TMS and direct cortical stimulation is similar to the logic behind aphasiology: if a particular language function is impaired when a specific region of the brain is knocked out, then that region must be somehow implicated in that language function. Few neurolinguistic studies to date have used TMS; direct cortical stimulation and cortical recording (recording brain activity using electrodes placed directly on the brain) have been used with macaque monkeys to make predictions about the behavior of human brains.

# Subject tasks

In many neurolinguistics experiments, subjects do not simply sit and listen to or watch stimuli, but also are instructed to perform some sort of task in response to the stimuli. Subjects perform these tasks while recordings (electrophysiological or

hemodynamic) are being taken, usually in order to ensure that they are paying attention to the stimuli. At least one study has suggested that the task the subject does has an effect on the brain responses and the results of the experiment.

## Lexical decision

The lexical decision task involves subjects seeing or hearing an isolated word and answering whether or not it is a real word. It is frequently used in priming studies, since subjects are known to make a lexical decision more quickly if a word has been primed by a related word (as in "doctor" priming "nurse").

## Grammaticality judgment, acceptability judgment

Many studies, especially violation-based studies, have subjects make a decision about the "acceptability" (usually grammatical acceptability or semantic acceptability) of stimuli. Such a task is often used to "ensure that subjects [are] reading the sentences attentively and that they [distinguish] acceptable from unacceptable sentences in the way the [experimenter] expect[s] them to do."

Experimental evidence has shown that the instructions given to subjects in an acceptability judgment task can influence the subjects' brain responses to stimuli. One experiment showed that when subjects were instructed to judge the "acceptability" of sentences they did not show an N400 brain response (a

response commonly associated with semantic processing), but that they did show that response when instructed to ignore grammatical acceptability and only judge whether or not the sentences "made sense".

## Probe verification

Some studies use a "probe verification" task rather than an overt acceptability judgment; in this paradigm, each experimental sentence is followed by a "probe word", and subjects must answer whether or not the probe word had appeared in the sentence. This task, like the acceptability judgment task, ensures that subjects are reading or listening attentively, but may avoid some of the additional processing demands of acceptability judgments, and may be used no matter what type of violation is being presented in the study.

## Truth-value judgment

Subjects may be instructed not to judge whether or not the sentence is grammatically acceptable or logical, but whether the proposition expressed by the sentence is true or false. This task is commonly used in psycholinguistic studies of child language.

## Active distraction and double-task

Some experiments give subjects a "distractor" task to ensure that subjects are not consciously paying attention to the

experimental stimuli; this may be done to test whether a certain computation in the brain is carried out automatically, regardless of whether the subject devotes attentional resources to it. For example, one study had subjects listen to non-linguistic tones (long beeps and buzzes) in one ear and speech in the other ear, and instructed subjects to press a button when they perceived a change in the tone; this supposedly caused subjects not to pay explicit attention to grammatical violations in the speech stimuli. The subjects showed a mismatch response (MMN) anyway, suggesting that the processing of the grammatical errors was happening automatically, regardless of attention—or at least that subjects were unable to consciously separate their attention from the speech stimuli.

Another related form of experiment is the double-task experiment, in which a subject must perform an extra task (such as sequential finger-tapping or articulating nonsense syllables) while responding to linguistic stimuli; this kind of experiment has been used to investigate the use of working memory in language processing.

# History

**Linguistics** is the scientific study of language. It involves an analysis of language form, language meaning, and language in context.

Linguistics began to be studied systematically by the Indian scholar Pānini in the 6th century BCE. Beginning around the 4th century BCE, Warring States period China also developed

its own grammatical traditions. Aristotle laid the foundation of Western linguistics as part of the study of rhetoric in his *Poetics* ca. 335 BC. Traditions of Arabic grammar and Hebrew grammar developed during the Middle Ages in a religious context like Pānini's Sanskrit grammar.

Modern approaches began to develop in the 18th century when the classical discipline of rhetoric was gradually removed. During the 19th century, linguistics came to be regarded as belonging to either psychology or biology, and such views remain the foundation of today's mainstream Anglo-American linguistics. They were however contested in the early 20th century by Ferdinand de Saussure who established linguistics as an autonomous discipline within social sciences. Following Saussure's concept, general linguistics consists of the study of language as a semioticsystem which includes the subfields of phonology, morphology, syntax, and semantics. The linguist's approach to these can be synchronic or diachronic.

Today, linguistics is considered as relating to a large number of scientific approaches and is further split into several subfields including applied linguistics, psycholinguistics, sociolinguistics, computational linguistics, and so on.

# Antiquity

Across cultures, the early history of linguistics is associated with a need to disambiguate discourse, especially for ritual texts or arguments. This often led to explorations of sound-meaning mappings, and the debate over conventional versus

naturalistic origins for these symbols. Finally, this led to the processes by which larger structures are formed from units.

## Babylonia

The earliest linguistic texts – written in cuneiform on clay tablets – date almost four thousand years before the present. In the early centuries of the second millennium BCE, in southern Mesopotamia, there arose a grammatical tradition that lasted more than 2,500 years. The linguistic texts from the earliest parts of the tradition were lists of nouns in Sumerian (a language isolate, that is, a language with no known genetic relatives), the language of religious and legal texts at the time. Sumerian was being replaced in everyday speech by a very different (and unrelated) language, Akkadian; it remained however as a language of prestige and continued to be used in religious and legal contexts. It therefore had to be taught as a foreign language, and to facilitate this, information about Sumerian was recorded in writing by Akkadian-speaking scribes.

Over the centuries, the lists became standardised, and the Sumerian words were provided with Akkadian translations. Ultimately texts emerged that gave Akkadian equivalents for not just single words, but for entire paradigms of varying forms for words: one text, for instance, has 227 different forms of the verb *ĝar* “to place”.

# India

Linguistics in ancient India derives its impetus from the need to correctly recite and interpret the Vedic texts. Already in the oldest Indian text, the Rigveda, *vāk* ("speech") is deified. By 1200 BCE, the oral performance of these texts becomes standardized, and treatises on ritual recitation suggest splitting up the Sanskrit compounds into words, stems, and phonetic units, providing an impetus for morphology and phonetics.

Some of the earliest activities in the description of language have been attributed to the Indian grammarian Pāāini (6th century BCE), who wrote a rule-based description of the Sanskrit language in his *Aĝĝādhyāyī*.

Over the next few centuries, clarity was reached in the organization of sound units, and the stop consonants were organized in a 5x5 square (c. 800 BCE, Pratisakhyas), eventually leading to a systematic alphabet, Brāhmī, by the 3rd century BCE.

In semantics, the early Sanskrit grammarian Śākaṭāyana (before c. 500 BCE) proposes that verbs represent ontologically prior categories, and that all nouns are etymologically derived from actions. The etymologist Yāska (c. 5th century BCE) posits that meaning inheres in the sentence, and that word meanings are derived based on sentential usage. He also provides four categories of words—nouns, verbs, pre-verbs, and particles/invariants—and a test for nouns both concrete and abstract: words which can be indicated by the pronoun *that*.

Pāṇini (c. 6th century BCE) opposes the Yāska view that sentences are primary, and proposes a grammar for composing semantics from morphemic roots. Transcending the ritual text to consider living language, Pāṇini specifies a comprehensive set of about 4,000 aphoristic rules (*sutras*) that:

- Map the semantics of verb argument structures into thematic roles
- Provide morphosyntactic rules for creating verb forms and nominal forms whose seven cases are called *karaka* (similar to case) that generate the morphology
- Take these morphological structures and consider phonological processes (e.g., root or stem modification) by which the final phonological form is obtained

In addition, the Pāā inian school also provides a list of 2000 verb roots which form the objects on which these rules are applied, a list of sounds (the so-called *Shiva-sutras*), and a list of 260 words not derivable by the rules.

The extremely succinct specification of these rules and their complex interactions led to considerable commentary and extrapolation over the following centuries. The phonological structure includes defining a notion of sound universals similar to the modern phoneme, the systematization of consonants based on oral cavity constriction, and vowels based on height and duration. However, it is the ambition of mapping these from morpheme to semantics that is truly remarkable in modern terms.

Grammarians following Pāṇini include Kātyāyana (c. 3rd century BCE), who wrote aphorisms on Pāṇini (the *Varttika*) and advanced mathematics; Patañjali (2nd century BCE), known for his commentary on selected topics in Pāṇini's grammar (the *Mahabhasya*) and on Kātyāyana'saphorisms, as well as, according to some, the author of the *Yoga Sutras*, and *Pingala*, with his mathematical approach to prosody. Several debates ranged over centuries, for example, on whether word-meaning mappings were conventional (*Vaisheshika-Nyaya*) or eternal (*Kātyāyana-Patañjali-Mīmāṃsā*).

The *Nyaya Sutras* specified three types of meaning: the individual (*this cow*), the type universal (*cowhood*), and the image (*draw the cow*). That the sound of a word also forms a class (sound-universal) was observed by Bhartṛhari (c. 500 CE), who also posits that language-universals are the units of thought, close to the nominalist or even the linguistic determinism position. Bhartṛhari also considers the sentence to be ontologically primary (word meanings are learned given their sentential use).

Of the six canonical texts or *Vedangas* that formed the core syllabus in Brahminic education from the 1st century CE until the 18th century, four dealt with language:

- *Shiksha* (*śikṣā*): phonetics and phonology (sandhi), Gārgeya and commentators
- *Chandas* (*chandas*): prosody or meter, Pingala and commentators
- *Vyakarana* (*vyākaraṇa*): grammar, Pāṇini and commentators

- *Nirukta* (*nirukta*): etymology, Yāska and commentators

Bhartrihari around 500 CE introduced a philosophy of meaning with his *sphoṭa* doctrine.

Unfortunately, Pāṇini's rule-based method of linguistic analysis and description has remained relatively unknown to Western linguistics until more recently. Franz Bopp used Pāṇini's work as a linguistic source for his 1807 Sanskrit grammar but disregarded his methodology. Pāṇini's system also differs from modern formal linguistics in that, since Sanskrit is a free word-order language, it did not provide syntactic rules. Formal linguistics, as first proposed by Louis Hjelmslev in 1943, is nonetheless based on the same concept that the expression of meaning is organised on different layers of linguistic form (including phonology and morphology).

The Pali Grammar of Kacchayana, dated to the early centuries CE, describes the language of the Buddhist canon.

## Greece

The Greeks developed an alphabet using symbols from the Phoenicians, adding signs for vowels and for extra consonants appropriate to their idiom (see Robins, 1997). In the Phoenicians and in earlier Greek writing systems, such as Linear B, graphemes indicated syllables, that is sound combinations of a consonant and a vowel. The addition of vowels by the Greeks was a major breakthrough as it facilitated the writing of Greek by representing both vowels and consonants with distinct graphemes. As a result of the

introduction of writing, poetry such as the Homeric poems became written and several editions were created and commented on, forming the basis of philology and criticism.

Along with written speech, the Greeks commenced studying grammatical and philosophical issues. A philosophical discussion about the nature and origins of language can be found as early as the works of Plato. A subject of concern was whether language was man-made, a social artifact, or supernatural in origin. Plato in his *Cratylus* presents the naturalisticview, that word meanings emerge from a natural process, independent of the language user. His arguments are partly based on examples of compounding, where the meaning of the whole is usually related to the constituents, although by the end he admits a small role for convention. The sophists and Socrates introduced dialectics as a new text genre. The Platonic dialogs contain definitions of the meters of the poems and tragedy, the form and the structure of those texts (see the *Republic* and *Phaidros*, *Ion*, etc.).

Aristotle supports the conventional origins of meaning. He defined the logic of speech and of the argument. Furthermore, Aristotle's works on rhetoricand poetics became of the utmost importance for the understanding of tragedy, poetry, public discussions etc. as text genres. Aristotle's work on logic interrelates with his special interest in language, and his work on this area was fundamentally important for the development of the study of language (*logos* in Greek means both "language" and "logic reasoning"). In *Categories*, Aristotle defines what is meant by "synonymous" or univocal words, what is meant by "homonymous" or equivocal words, and what is meant by

"paronymous" or denominative words. He divides forms of speech as being:

- Either simple, without composition or structure, such as "man,""horse," "fights," etc.
- Or having composition and structure, such as "a man fights," "the horse runs," etc.

Next, he distinguishes between a subject of predication, namely that of which anything is affirmed or denied, and a subject of inhesion. A thing is said to be inherent in a subject, when, though it is not a part of the subject, it cannot possibly exist without the subject, e.g., shape in a thing having a shape. The categories are not abstract platonic entities but are found in speech, these are substance, quantity, quality, relation, place, time, position, state, action and affection. In *de Interpretatione*, Aristotle analyzes categoric propositions, and draws a series of basic conclusions on the routine issues of classifying and defining basic linguistic forms, such as simple terms and propositions, nouns and verbs, negation, the quantity of simple propositions (primitive roots of the quantifiers in modern symbolic logic), investigations on the excluded middle (which to Aristotle isn't applicable to future tense propositions — the Problem of future contingents), and on modal propositions.

The Stoics made linguistics an important part of their system of the cosmos and the human. They played an important role in defining the linguistic sign-terms adopted later on by Ferdinand de Saussure like "significant" and "signifié". The Stoics studied phonetics, grammar and etymology as separate levels of study. In phonetics and phonology the articulators

were defined. The syllable became an important structure for the understanding of speech organization. One of the most important contributions of the Stoics in language study was the gradual definition of the terminology and theory echoed in modern linguistics.

Alexandrian grammarians also studied speech sounds and prosody; they defined parts of speech with notions such as "noun", "verb", etc. There was also a discussion about the role of analogy in language, in this discussion the grammatici in Alexandria supported the view that language and especially morphology is based on analogy or paradigm, whereas the grammatic in schools in Asia Minor consider that language is not based on analogical bases but rather on exceptions.

Alexandrians, like their predecessors, were very interested in meter and its role in poetry. The metrical "feet" in the Greek was based on the length of time taken to pronounce each syllable, with syllables categorized according to their weight as either "long" syllables or "short" syllables (also known as "heavy" and "light" syllables, respectively, to distinguish them from long and short vowels). The foot is often compared to a musical measure and the long and short syllables to whole notes and half notes. The basic unit in Greek and Latin prosody is a mora, which is defined as a single short syllable. A long syllable is equivalent to two moras. A long syllable contains either a long vowel, a diphthong, or a short vowel followed by two or more consonants.

Various rules of elision sometimes prevent a grammatical syllable from making a full syllable, and certain other lengthening and shortening rules (such as correption) can

create long or short syllables in contexts where one would expect the opposite. The most important Classical meter as defined by the Alexandrian grammarians was the dactylic hexameter, the meter of Homeric poetry. This form uses verses of six feet. The first four feet are normally dactyls, but can be spondees. The fifth foot is almost always a dactyl. The sixth foot is either a spondee or a trochee. The initial syllable of either foot is called the ictus, the basic "beat" of the verse. There is usually a caesura after the ictus of the third foot.

The text *Tékhnēgrammatiké* (c. 100 BCE, Gk. *gramma* meant letter, and this title means "Art of letters"), possibly written by Dionysius Thrax (170 – 90 BCE), is considered the earliest grammar book in the Greek tradition. It lists eight parts of speech and lays out the broad details of Greek morphology including the case structures. This text was intended as a pedagogic guide (as was Panini), and also covers punctuation and some aspects of prosody. Other grammars by Charisius (mainly a compilation of Thrax, as well as lost texts by RemmiusPalaemon and others) and Diomedes (focusing more on prosody) were popular in Rome as pedagogic material for teaching Greek to native Latin-speakers.

One of the most prominent scholars of Alexandria and of the antiquity was Apollonius Dyscolus. Apollonius wrote more than thirty treatises on questions of syntax, semantics, morphology, prosody, orthography, dialectology, and more. Happily, four of these are preserved—we still have a *Syntax* in four books, and three one-book monographs on pronouns, adverbs, and connectives, respectively.

Lexicography become an important domain of study as many grammarians compiled dictionaries, thesauri and lists of special words "λέξεις" that were old, or dialectical or special (such as medical words or botanic words) at that period. In the early medieval times we find more categories of dictionaries like the dictionary of Suida (considered the first encyclopedic dictionary), etymological dictionaries etc.

At that period, the Greek language functioned as a *lingua franca*, a language spoken throughout the known world (for the Greeks and Romans) of that time and, as a result, modern linguistics struggles to overcome this. With the Greeks a tradition commenced in the study of language. The Romans and the medieval world followed, and their laborious work is considered today as a part of our everyday language. Think, for example, of notions such as the word, the syllable, the verb, the subject etc.

## Rome

In the 4th century, AeliusDonatus compiled the Latin grammar *ArsGrammatica* that was to be the defining school text through the Middle Ages. A smaller version, *Ars Minor*, covered only the eight parts of speech; eventually when books came to be printed in the 15th century, this was one of the first books to be printed. Schoolboys subjected to all this education gave us the current meaning of "grammar" (attested in English since 1176).

# China

Similar to the Indian tradition, Chinese philology, *Xiaoxue* (小學 "elementary studies"), began as an aid to understanding classics in the Han dynasty (c. 3rd century BCE). *Xiaoxue* came to be divided into three branches: *Xungu* (訓詁 "exegesis"), *Wenzi* (文字 "script [analysis]") and *Yinyun* (音韻 "[study of] sounds") and reached its golden age in the 17th century CE (Qing Dynasty). The glossary *Erya* (c. 3rd century BCE), comparable to the Indian *Nighantu*, is regarded as the first linguistic work in China. *ShuowenJiezi* (c. 2nd century BCE), the first Chinese dictionary, classifies Chinese characters by radicals, a practice that would be followed by most subsequent lexicographers. Two more pioneering works produced during the Han Dynasty are *Fangyan*, the first Chinese work concerning dialects, and *Shiming*, devoted to etymology.

As in ancient Greece, early Chinese thinkers were concerned with the relationship between names and reality. Confucius (6th century BCE) famously emphasized the moral commitment implicit in a name, (*zhengming*) stating that the moral collapse of the pre-Qin was a result of the failure to rectify behaviour to meet the moral commitment inherent in names: "Good government consists in the ruler being a ruler, the minister being a minister, the father being a father, and the son being a son... If names be not correct, language is not in accordance with the truth of things." (*Analects* 12.11,13.3).

However, what is the reality implied by a name? The later Mohists or the group known as School of Names (mingjia, 479-221 BCE), consider that *ming* (名 "name") may refer to three

kinds of *shi* (實 "actuality"): type universals (horse), individual (John), and unrestricted (thing). They adopt a realist position on the name-reality connection - universals arise because "the world itself fixes the patterns of similarity and difference by which things should be divided into kinds". The philosophical tradition is well known for conundra resembling the sophists, e.g. when GongsunLongzi (4th century BCE) questions if in copula statements (*X is Y*), are *X* and *Y* identical or is *X* a subclass of *Y*. This is the famous paradox "a white horse is not a horse".

XunZi (3rd century BCE) revisits the principle of *zhengming*, but instead of rectifying behaviour to suit the names, his emphasis is on rectifying language to correctly reflect reality. This is consistent with a more "conventional" view of word origins (*yuedingsucheng*約定俗成).

The study of phonology in China began late, and was influenced by the Indian tradition, after Buddhism had become popular in China. The rime dictionary is a type of dictionary arranged by tone and rime, in which the pronunciations of characters are indicated by *fanqie* spellings. Rime tableswere later produced to aid the understanding of *fanqie*.

Philological studies flourished during the Qing Dynasty, with DuanYucai and Wang Niansun as the towering figures. The last great philologist of the era was Zhang Binglin, who also helped lay the foundation of modern Chinese linguistics. The Western comparative methodwas brought into China by Bernard Karlgren, the first scholar to reconstruct Middle Chinese and Old Chinese with Latin alphabet (not IPA). Important modern

Chinese linguists include Y. R. Chao, Luo Changpei, Li Fanggui and Wang Li.

The ancient commentators on the classics paid much attention to syntax and the use of particles. But the first Chinese grammar, in the modern sense of the word, was produced by Ma Jianzhong (late 19th century). His grammar was based on the Latin (prescriptive) model.

# Middle Ages

## Arabic grammar

Owing to the rapid expansion of Islam in the 8th century, many people learned Arabic as a lingua franca. For this reason, the earliest grammatical treatises on Arabic are often written by non-native speakers.

The earliest grammarian who is known to us is *ʿAbdAllāh ibn AbīIsḥāq al-Ḥaḍramī* (died 735-736 CE, 117 AH). The efforts of three generations of grammarians culminated in the book of the Persian linguist *Sibāwayhi* (c. 760–793).

Sibawayh made a detailed and professional description of Arabic in 760 in his monumental work, *Al-kitab fi al-nahw* (الكتابفيالنحو, *The Book on Grammar*). In his book he distinguished phonetics from phonology.

# European vernaculars

The Irish *SanasCormaic* 'Cormac's Glossary' is Europe's first etymological and encyclopedic dictionary in any non-Classical language.

The Modistae or "speculative grammarians" in the 13th century introduced the notion of universal grammar.

In *De vulgarieloquentia* ("On the Eloquence of Vernacular"), Dante expanded the scope of linguistic enquiry from Latin/Greek to include the languages of the day. Other linguistic works of the same period concerning the vernaculars include the First Grammatical Treatise (Icelandic) or the Auraiceptna n-Éces (Irish).

The Renaissance and Baroque period saw an intensified interest in linguistics, notably for the purpose of Bible translations by the Jesuits, and also related to philosophical speculation on philosophical languages and the origin of language.

Founding Fathers In the 1600s, JoannesGoropiusBecanus was the oldest representative of Dutch linguistics. He was the first person to publish a fragment of Gothic, mainly The lord's prayer. FranciscusJuniuns, Lambert ten Kate from Amsterdam and George Hickes from England are considered to be the founding fathers of German linguistics.

# Modern linguistics

Modern linguistics did not begin until the late 18th century, and the Romantic or animist theses of Johann Gottfried Herder and Johann Christoph Adelung remained influential well into the 19th century.

In the history of American linguistics, there were hundreds of Indigenous languages that were never recorded. Many of the languages were spoken and so they are now inaccessible. Under these circumstances, linguistics such as Franz Boas tried to prescribed sound methodical principles for the analysis of unfamiliar languages. Boas was an influential linguist and was followed by Edward Sapir and Leonard Bloomfield.

## Historical linguistics

During the 18th century conjectural history, based on a mix of linguistics and anthropology, on the topic of both the origin and progress of language and society was fashionable. These thinkers contributed to the construction of academic paradigms in which some languages were labelled "primitive" relative to the English language. Within this paradigm a primitive people could be discerned by their primitive language, as in the case of Hugh Blair who argued that Native Americans gesticulated wildly to compensate for poor lexicon of their primitive language. Around the same time, James Burnett authored a 6 volume treatise that delved more deeply into the matter of "savage languages". Other writers theorized that Native American languages were "nothing but the natural and instinctive cries of the animal" without grammatical structure.

The thinkers within this paradigm connected themselves with the Greeks and Romans, viewed as the only civilized persons of the ancient world, a view articulated by Thomas Sheridan who compiled an important 18th century pronunciation dictionary: "It was to the care taken in the cultivation of their languages, that Greece and Rome, owed that splendor, which eclipsed all the other nations of the world".

In the 18th century James Burnett, Lord Monboddo analyzed numerous languages and deduced logical elements of the evolution of human languages. His thinking was interleaved with his precursive concepts of biological evolution. Some of his early concepts have been validated and are considered correct today. In his *The Sanscrit Language* (1786), Sir William Jones proposed that Sanskrit and Persian had resemblances to Classical Greek, Latin, Gothic, and Celtic languages. From this idea sprung the field of comparative linguistics and historical linguistics. Through the 19th century, European linguistics centered on the comparative history of the Indo-European languages, with a concern for finding their common roots and tracing their development.

In 1786, it was discovered that there is a regular sound that corresponded in the languages spoken in Europe, India, and Persia. This led to the conclusion that all of the languages can from a common ancestor and during the 19th-century linguistics were devoted to figuring out the nuances of the parent language. It was discovered that this parent language started approximately 6000 years ago and has also developed in English, Russian, and Hindi.

In the 1820s, Wilhelm von Humboldt observed that human language was a rule-governed system, anticipating a theme that was to become central in the formal work on syntax and semantics of language in the 20th century. Of this observation he said that it allowed language to make "infinite use of finite means" (*Über den Dualis*, 1827). Humboldt's work is associated with the movement of **Romantic linguistics**, which was inspired by *Naturphilosophie* and Romantic science. Other notable representatives of the movement include Friedrich Schlegel and Franz Bopp.

It was only in the late 19th century that the Neogrammarian approach of Karl Brugmann and others introduced a rigid notion of sound law.

Historical linguistics also led to the emergence of the semantics and some forms of pragmatics (Nerlich, 1992; Nerlich and Clarke, 1996).

Historical linguistics continues today and linguistics have succeeded in grouping approximately 5000 languages of the world into a number of common ancestors.

## Structuralism

In Europe there was a development of structural linguistics, initiated by Ferdinand de Saussure, a Swiss professor of Indo-European and general linguistics, whose lectures on general linguistics, published posthumously by his students, set the direction of European linguistic analysis from the 1920s on; his approach has been widely adopted in other fields under the broad term "Structuralism".

By the 20th century, the attention shifted from language change to the structure, which is governed by rules and principles. This structure turned more into grammar and by the 1920s structural linguistic, was developing into sophisticated methods of grammatical analysis.

## Descriptive linguistics

During the second World War, North American linguists Leonard Bloomfield, William Mandeville Austin and several of his students and colleagues developed teaching materials for a variety of languages whose knowledge was needed for the war effort. This work led to an increasing prominence of the field of linguistics, which became a recognized discipline in most American universities only after the war.

In 1965, William Stokoe, a linguist from Gallaudet University published an analysis [1] which proved that American Sign Language fits the criteria for a natural language.

## Other subfields

From roughly 1980 onwards, pragmatic, functional, and cognitive approaches have steadily gained ground, both in the United States and in Europe.

# Language documentation

**Language documentation** (also:**documentary linguistics**) is a subfield of linguistics which aims to describe the grammar and

use of human languages. It aims to provide a comprehensive record of the linguistic practices characteristic of a given speech community. Language documentation seeks to create as thorough a record as possible of the speech community for both posterity and language revitalization. This record can be public or private depending on the needs of the community and the purpose of the documentation. In practice, language documentation can range from solo linguistic anthropological fieldwork to the creation of vast online archives that contain dozens of different languages, such as FirstVoices or OLAC.

Language documentation provides a firmer foundation for linguistic analysis in that it creates a corpus of materials in the language. The materials in question can range from vocabulary lists and grammar rules to children's books and translated works. These materials can then support claims about the structure of the language and its usage.

# Methods

Typical steps involve recording, maintaining metadata, transcribing (often using the International Phonetic Alphabet and/or a "practical orthography" made up for that language), annotation and analysis, translation into a language of wider communication, archiving and dissemination. Critical is the creation of good records in the course of doing language description. The materials can be archived, but not all archives are equally adept at handling language materials preserved in varying technological formats, and not all are equally accessible to potential users.

Language documentation complements language description, which aims to describe a language's abstract system of structures and rules in the form of a grammar or dictionary. By practicing good documentation in the form of recordings with transcripts and then collections of texts and a dictionary, a linguist works better and can provide materials for use by speakers of the language. New technologies permit better recordings with better descriptions which can be housed in digital archives such as AILLA or PARADISEC. These resources can then be made available to the speakers.

Language documentation has also given birth to new specialized publications, such as the free online and peer-reviewed journals *Language Documentation & Description* and *Language Documentation & Conservation.*

# Digital language archives

The digitization of archives is a critical component of language documentation and revitalization projects. There are descriptive records of local languages that could be put to use in language revitalization projects that are overlooked due to obsolete formatting, incomplete hard-copy records, or systematic inaccessibility. Local archives in particular, which may have vital records of the area's indigenous languages, are chronically underfunded and understaffed. Historic records relating to language that have been collected by non-linguists such as missionaries can be overlooked if the collection is not digitized. Physical archives are naturally more vulnerable to damage and information loss.

# Teaching with documentation

Language documentation can be beneficial to individuals who would like to teach or learn an endangered language. If a language has limited documentation this also limits how it can be used in a language revitalization context. Teaching with documentation and linguist's field notes can provide more context for those teaching the language and can add information they were not aware of. Documentation can be useful for understanding culture and heritage, as well as learning the language. Important components when teaching a language includes: Listening, reading, speaking, writing, and cultural components. Documentation gives resources to further the skills for learning a language. For example, the Kaurna language was revitalized through written resources. These written documents served as the only resource and were used to re-introduce the language and one way was through teaching, which also included the making of a teaching guide for the Kaurna language. Language documentation and teaching have a relationship because if there are no fluent speakers of a language, documentation can be used as a teaching resource.

# Types

Language description, as a task within linguistics, may be divided into separate areas of specialization:

- Phonetics, the study of the sounds of human language

- Phonology, the study of the sound system of a language
- Morphology, the study of the internal structure of words
- Syntax, the study of how words combine to form grammatical sentences
- Semantics, the study of the meaning of words (lexical semantics), and how these combine to form the meanings of sentences
- Historical linguistics, the study of languages whose historical relations are recognizable through similarities in vocabulary, word formation, and syntax
- Pragmatics, the study of how language is used by its speakers
- Stylistics, the study of style in languages
- Paremiography, the collection of proverbs and sayings

# Related research areas

- Linguistic description
- Orthography, the study of writing systems
- Lexicography, the study and practice of making dictionaries
- Phonology, the study of describing the sound system of a language
- Etymology, the study of how words acquire their meanings
- Anthropological linguistics

# Organizations

- HRELP
- DoBeS
- First Peoples' Heritage, Language and Culture Council
- LACITO and the Pangloss Collection
- The Language Conservancy
- PARADISEC Archive
- The Endangered Languages Archive (ELAR)
- Resource Network for Linguistic Diversity
- SIL International
- Western Institute for Endangered Language Documentation (WIELD)
- World Oral Literature Project, Voices of Vanishing Worlds

Chapter 4

# Corpus and Translation Studies

## Corpus linguistics

**Corpus linguistics** is the study of a language as that language is expressed in its text corpus (plural *corpora*), its body of "real world" text. Corpus linguistics proposes that a reliable analysis of a language is more feasible with a corpora collected in the field—the natural context ("realia") of that language—with minimal experimental interference.

The text-corpus method uses the body of texts written in any natural language to derive the set of abstract rules which govern that language. Those results can be used to explore the relationships between that subject language and other languages which have undergone a similar analysis. The first such corpora were manually derived from source texts, but now that work is automated.

Corpora have not only been used for linguistics research, they have also been used to compile dictionaries (starting with *The American Heritage Dictionary of the English Language* in 1969) and grammar guides, such as *A Comprehensive Grammar of the English Language*, published in 1985.

Experts in the field have differing views about the annotation of a corpus. These views range from John McHardy Sinclair, who advocates minimal annotation so texts speak for themselves, to the Survey of English Usage team (University

College, London), who advocate annotation as allowing greater linguistic understanding through rigorous recording.

# History

Some of the earliest efforts at grammatical description were based at least in part on corpora of particular religious or cultural significance. For example, Prātiśākhya literature described the sound patterns of Sanskrit as found in the Vedas, and Pāṇini's grammar of classical Sanskritwas based at least in part on analysis of that same corpus. Similarly, the early Arabic grammarians paid particular attention to the language of the Qur̲an. In the Western European tradition, scholars prepared concordances to allow detailed study of the language of the Bible and other canonical texts.

## English corpora

A landmark in modern corpus linguistics was the publication of *Computational Analysis of Present-Day American English* in 1967. Written by Henry Kučera and W. Nelson Francis, the work was based on an analysis of the Brown Corpus, which was a contemporary compilation of about a million American English words, carefully selected from a wide variety of sources. Kučera and Francis subjected the Brown Corpus to a variety of computational analyses and then combined elements of linguistics, language teaching, psychology, statistics, and sociology to create a rich and variegated opus. A further key publication was Randolph Qu̲irk's "Towards a description of

English Usage" in 1960 in which he introduced the Survey of English Usage.

Shortly thereafter, Boston publisher Houghton-Mifflin approached Kučera to supply a million-word, three-line citation base for its new *American Heritage Dictionary*, the first dictionary compiled using corpus linguistics. The *AHD* took the innovative step of combining prescriptive elements (how language *should* be used) with descriptive information (how it actually *is* used).

Other publishers followed suit. The British publisher Collins' COBUILDmonolingual learner's dictionary, designed for users learning English as a foreign language, was compiled using the Bank of English. The Survey of English Usage Corpus was used in the development of one of the most important Corpus-based Grammars, which was written by Quirk *et al.* and published in 1985 as *A Comprehensive Grammar of the English Language*.

The Brown Corpus has also spawned a number of similarly structured corpora: the LOB Corpus (1960s British English), Kolhapur (Indian English), Wellington (New Zealand English), Australian Corpus of English (Australian English), the Frown Corpus (early 1990sAmerican English), and the FLOB Corpus (1990s British English). Other corpora represent many languages, varieties and modes, and include the International Corpus of English, and the British National Corpus, a 100 million word collection of a range of spoken and written texts, created in the 1990s by a consortium of publishers, universities (Oxford and Lancaster) and the British Library. For contemporary American English, work has stalled on the American National Corpus, but the 400+ million word Corpus

of Contemporary American English (1990–present) is now available through a web interface.

The first computerized corpus of transcribed spoken language was constructed in 1971 by the Montreal French Project, containing one million words, which inspired Shana Poplack's much larger corpus of spoken French in the Ottawa-Hull area.

## Multilingual Corpora

In the 1990s, many of the notable early successes on statistical methods in natural-language programming (NLP) occurred in the field of machine translation, due especially to work at IBM Research. These systems were able to take advantage of existing multilingual textual corpora that had been produced by the Parliament of Canada and the European Unionas a result of laws calling for the translation of all governmental proceedings into all official languages of the corresponding systems of government.

There are corpora in non-European languages as well. For example, the National Institute for Japanese Language and Linguistics in Japan has built a number of corpora of spoken and written Japanese.

## Ancient languages corpora

Besides these corpora of living languages, computerized corpora have also been made of collections of texts in ancient languages. An example is the Andersen-Forbes database of the Hebrew Bible, developed since the 1970s, in which every clause

is parsed using graphs representing up to seven levels of syntax, and every segment tagged with seven fields of information. The Quranic Arabic Corpus is an annotated corpus for the Classical Arabic language of the Quran. This is a recent project with multiple layers of annotation including morphological segmentation, part-of-speech tagging, and syntactic analysis using dependency grammar.

## Corpora from specific fields

Besides pure linguistic inquiry, researchers had begun to apply corpus linguistics to other academic and professional fields, such as the emerging sub-discipline of law and corpus linguistics, which seeks to understand legal texts using corpus data and tools.

# Methods

Corpus linguistics has generated a number of research methods, which attempt to trace a path from data to theory. Wallis and Nelson (2001) first introduced what they called the 3A perspective: Annotation, Abstraction and Analysis.

- **Annotation** consists of the application of a scheme to texts. Annotations may include structural markup, part-of-speech tagging, parsing, and numerous other representations.
- **Abstraction** consists of the translation (mapping) of terms in the scheme to terms in a theoretically motivated model or dataset. Abstraction typically

includes linguist-directed search but may include e.g., rule-learning for parsers.

- **Analysis** consists of statistically probing, manipulating and generalising from the dataset. Analysis might include statistical evaluations, optimisation of rule-bases or knowledge discovery methods.

Most lexical corpora today are part-of-speech-tagged (POS-tagged). However even corpus linguists who work with 'unannotated plain text' inevitably apply some method to isolate salient terms. In such situations annotation and abstraction are combined in a lexical search.

The advantage of publishing an annotated corpus is that other users can then perform experiments on the corpus (through corpus managers). Linguists with other interests and differing perspectives than the originators' can exploit this work. By sharing data, corpus linguists are able to treat the corpus as a locus of linguistic debate and further study.

# Translation studies

**Translation studies** is an academic interdiscipline dealing with the systematic study of the theory, description and application of translation, interpreting, and localization. As an interdiscipline, translation studies borrows much from the various fields of study that support translation. These include comparative literature, computer science, history, linguistics, philology, philosophy, semiotics, and terminology.

The term "translation studies" was coined by the Amsterdam-based American scholar James S. Holmes in his paper "The name and nature of translation studies", which is considered a foundational statement for the discipline. English writers, occasionally use the term "**translatology**" (and less commonly "**traductology**") to refer to translation studies, and the corresponding French term for the discipline is usually "*traductologie*" (as in the SociétéFrançaise de Traductologie). In the United States, there is a preference for the term "translation and interpreting studies" (as in the American Translation and Interpreting Studies Association), although European tradition includes interpreting within translation studies (as in the European Society for Translation Studies).

# History

## Early studies

Historically, translation studies has long been "prescriptive" (telling translators how to translate), to the point that discussions of translation that were not prescriptive were generally not considered to be about translation at all. When historians of translation studies trace early Western thought about translation, for example, they most often set the beginning at the renowned orator Cicero's remarks on how he used translation from Greek to Latin to improve his oratorical abilities — an early description of what Jerome ended up calling sense-for-sense translation. The descriptive history of interpreters in Egypt provided by Herodotus several centuries earlier is typically not thought of as translation studies —

presumably because it does not tell translators how to translate. In China, the discussion on how to translate originated with the translation of Buddhist sutras during the Han Dynasty.

## Calls for an academic discipline

In 1958, at the Fourth Congress of Slavists in Moscow, the debate between linguistic and literary approaches to translation reached a point where it was proposed that the best thing might be to have a separate science that was able to study all forms of translation, without being wholly within linguistics or wholly within literary studies. Within comparative literature, translation workshops were promoted in the 1960s in some American universities like the University of Iowa and Princeton.

During the 1950s and 1960s, systematic linguistic-oriented studies of translation began to appear. In 1958, the French linguists Jean-Paul Vinay and Jean Darbelnet carried out a contrastive comparison of French and English. In 1964, Eugene Nida published *Toward a Science of Translating*, a manual for Bible translation influenced to some extent by Harris's transformational grammar. In 1965, J. C. Catford theorized translation from a linguistic perspective. In the 1960s and early 1970s, the Czech scholar JiříLevý and the Slovak scholars Anton Popovič and FrantišekMiko worked on the stylistics of literary translation.

These initial steps toward research on literary translation were collected in James S. Holmes' paper at the Third International Congress of Applied Linguistics held in Copenhagen in 1972. In

that paper, "The name and nature of translation studies", Holmes asked for the consolidation of a separate discipline and proposed a classification of the field. A visual "map" of Holmes' proposal was later presented by Gideon Toury in his 1995 *Descriptive Translation Studies and beyond*.

Before the 1990s, translation scholars tended to form particular schools of thought, particularly within the prescriptive, descriptive and Skopos paradigms. Since the "cultural turn" in the 1990s, the discipline has tended to divide into separate fields of inquiry, where research projects run parallel to each other, borrowing methodologies from each other and from other academic disciplines.

# Schools of thought

The main schools of thought on the level of research have tended to cluster around key theoretical concepts, most of which have become objects of debate.

## Equivalence

Through to the 1950s and 1960s, discussions in translation studies tended to concern how best to attain "equivalence". The term "equivalence" had two distinct meanings, corresponding to different schools of thought. In the Russian tradition, "equivalence" was usually a one-to-one correspondence between linguistic forms, or a pair of authorized technical terms or phrases, such that "equivalence" was opposed to a range of "substitutions". However, in the French tradition of Vinay and Darbelnet, drawing on Bally, "equivalence" was the

attainment of equal functional value, generally requiring *changes* in form. Catford's notion of equivalence in 1965 was as in the French tradition. In the course of the 1970s, Russian theorists adopted the wider sense of "equivalence" as something *resulting* from linguistic transformations.

At about the same time, the *Interpretive Theory of Translation* introduced the notion of deverbalized sense into translation studies, drawing a distinction between word correspondences and sense equivalences, and showing the difference between dictionary definitions of words and phrases (word correspondences) and the sense of texts or fragments thereof in a given context (sense equivalences).

The discussions of equivalence accompanied typologies of translation solutions (also called "procedures", "techniques" or "strategies"), as in Fedorov (1953) and Vinay and Darbelnet (1958). In 1958, LohDianyang's*Translation: Its Principles and Techniques* (英汉翻译理论与技巧) drew on Fedorov and English linguistics to present a typology of translation solutions between Chinese and English.

In these traditions, discussions of the ways to attain equivalence have mostly been prescriptive and have been related to translator training.

## Descriptive translation studies

Descriptive translation studies aims at building an empirical descriptive discipline, to fill one section of the Holmes map. The idea that scientific methodology could be applicable to cultural products had been developed by the Russian

Formalists in the early years of the 20th century, and had been recovered by various researchers in comparative literature. It was now applied to literary translation. Part of this application was the theory of polysystems (Even-Zohar 1990) in which translated literature is seen as a sub-system of the receiving or target literary system. Gideon Toury bases his theory on the need to consider translations as "facts of the target culture" for the purposes of research. The concepts of "manipulation" and "patronage" have also been developed in relation to literary translations.

## Skopos theory

Another paradigm shift in translation theory can be dated from 1984 in Europe and the publication of two books in German: *Foundation for a General Theory of Translation* by Katharina Reiss (also written Reiß) and Hans Vermeer, and *Translatorial Action* (TranslatorischesHandeln) by JustaHolz-Mänttäri. From these two came what is known as Skopos theory, which gives priority to the purpose to be fulfilled by the translation instead of prioritizing equivalence.

## Cultural translation

The cultural turn meant still another step forward in the development of the discipline. It was sketched by Susan Bassnett and André Lefevere in *Translation - History - Culture*, and quickly represented by the exchanges between translation studies and other area studies and concepts: gender studies, cannibalism, post-colonialism or cultural studies, among others.

The concept of "cultural translation" largely ensues from HomiBhabha's reading of Salman Rushdie in *The Location of Culture*. Cultural translation is a concept used in cultural studies to denote the process of transformation, linguistic or otherwise, in a given culture. The concept uses linguistic translation as a tool or metaphor in analyzing the nature of transformation and interchange in cultures.

# Fields of inquiry

## Translation history

Translation history concerns the history of translators as a professional and social group, as well as the history of translations as indicators of the way cultures develop, interact and may die. Some principles for translation history have been proposed by LievenD'hulst and Pym. Major projects in translation history have included the *Oxford History of Literary Translation in English* and *Histoire des traductionsen langue française*.

Historical anthologies of translation theories have been compiled by Robinson (2002) for Western theories up to Nietzsche; by D'hulst (1990) for French theories, 1748–1847; by Santoyo (1987) for the Spanish tradition; by Edward Balcerzan (1977) for the Polish experience, 1440–1974; and by Cheung (2006) for Chinese.

## Sociologies of translation

The sociology of translation includes the study of who translators are, what their forms of work are (workplace studies) and what data on translations can say about the movements of ideas between languages.

## Post-colonial translation studies

Post-colonial studies look at translations between a metropolis and former colonies, or within complex former colonies. They radically question the assumption that translation occurs between cultures and languages that are radically separated.

## Gender studies

Gender studies look at the sexuality of translators, at the gendered nature of the texts they translate, at the possibly gendered translation processes employed, and at the gendered metaphors used to describe translation. Pioneering studies are by Luise von Flotow, Sherry Simon and Keith Harvey. The effacement or inability to efface threatening forms of same-sex sexuality is a topic taken up, when for instance ancient writers are translated by Renaissance thinkers in a Christian context.

## Ethics

In the field of ethics, much-discussed publications have been the essays of Antoine Berman and Lawrence Venuti that differ in some aspects but agree on the idea of emphasizing the

differences between source and target language and culture when translating. Both are interested in how the "cultural other [...] can best preserve [...] that otherness". In more recent studies, scholars have applied Emmanuel Levinas' philosophical work on ethics and subjectivity on this issue. As his publications have been interpreted in different ways, various conclusions on his concept of ethical responsibility have been drawn from this. Some have come to the assumption that the idea of translation itself could be ethically doubtful, while others receive it as a call for considering the relationship between author or text and translator as more interpersonal, thus making it an equal and reciprocal process.

Parallel to these studies, the general recognition of the translator's responsibility has increased. More and more translators and interpreters are being seen as active participants in geopolitical conflicts, which raises the question of how to act ethically independent from their own identity or judgement. This leads to the conclusion that translating and interpreting cannot be considered solely as a process of language transfer, but also as socially and politically directed activities.

There is general agreement on the need for an ethical code of practice providing some guiding principles to reduce uncertainties and improve professionalism, as having been stated in other disciplines (for example military medical ethics or legal ethics). However, as there is still no clear understanding of the concept of ethics in this field, opinions about the particular appearance of such a code vary considerably.

## Audiovisual translation studies

Audiovisual translation studies (AVT) is concerned with translation that takes place in audio and/or visual settings, such as the cinema, television, video games and also some live events such as opera performances. The common denominator for studies in this field is that translation is carried out on multiple semiotic systems, as the translated texts (so-called polysemiotic texts) have messages that are conveyed through more than one semiotic channel, i.e. not just through the written or spoken word, but also via sound and/or images. The main translation modes under study are subtitling, film dubbing and voice-over, but alsosurtitling for the opera and theatre.

Media accessibility studies is often considered a part of this field as well, with audio description for the blind and partially sighted and subtitles for the deaf or hard-of-hearing being the main objects of study. The various conditions and constraints imposed by the different media forms and translation modes, which influence how translation is carried out, are often at the heart of most studies of the product or process of AVT. Many researchers in the field of AVT Studies are organized in the European Association for Studies in Screen Translation, as are many practitioners in the field.

## Non-professional translation

Non-professional translation refers to the translation activities performed by translators who are not working professionally, usually in ways made possible by the Internet. These practices

have mushroomed with the recent democratization of technology and the popularization of the Internet. Volunteer translation initiatives have emerged all around the world, and deal with the translations of various types of written and multimedia products.

Normally, it is not required for volunteers to have been trained in translation, but trained translators could also participate, such as the case of Translators without Borders.

Depending on the feature that each scholar considers the most important, different terms have been used to label "non-professional translation". O'Hagan has used "user-generated translation", "fan translation" and "community translation". Fernández-Costales and Jiménez-Crespo prefer "collaborative translation", while Pérez-González labels it "amateur subtitling". Pym proposes that the fundamental difference between this type of translation and professional translation relies on monetary reward, and he suggests it should be called "volunteer translation".

Some of the most popular fan-controlled non-professional translation practices are fansubbing, fandubbing, ROM hacking or fan translation of video games, and scanlation. These practices are mostly supported by a strong and consolidated fan base, although larger non-professional translation projects normally apply crowdsourcing models and are controlled by companies or organizations. Since 2008, *Facebook* has used crowdsourcing to have its website translated by its users and TED conference has set up the open translation project TED Translators in which volunteers

use the Amara platform to create subtitles online for TED talks.

## Localization

Studies of localization concern the way the contemporary language industries translate and adapt ("localize") technical texts across languages, tailoring them for a specific "locale" (a target location defined by language variety and various cultural parameters). Localization usually concerns software, product documentation, websites and video games, where the technological component is key.

A key concept in localization is internationalization, in which the start product is stripped of its culture-specific features in such a way that it can be simultaneously localized into several languages.

## Translator education

## Interpreting Studies

The discipline of interpreting studies is often referred to as the sister of translation studies. This is due to the similarities between the two disciplines, consisting in the transfer of ideas from one language into another. Indeed, interpreting as an activity was long seen as a specialized form of translation, before scientifically founded interpreting studies emancipated gradually from translation studies in the second half of the 20th century. While they were strongly oriented towards the theoretic framework of translation studies, interpreting studies have always been concentrating on the practical and

pedagogical aspect of the activity. This led to the steady emancipation of the discipline and the consecutive development of a separate theoretical framework based - as are translation studies - on interdisciplinary premises. Interpreting studies have developed several approaches and undergone various paradigm shifts, leading to the most recent surge of sociological studies of interpreters and their (working conditions).

## Cognition and process studies

## Translation technologies

# Future prospects

Translation studies has developed alongside the growth in translation schools and courses at the university level. In 1995, a study of 60 countries revealed there were 250 bodies at university level offering courses in translation or interpreting. In 2013, the same database listed 501 translator-training institutions. Accordingly, there has been a growth in conferences on translation, translation journals and translation-related publications. The visibility acquired by translation has also led to the development of national and international associations of translation studies. Ten of these associations formed the International Network of Translation and Interpreting Studies Associations in September 2016.

The growing variety of paradigms is mentioned as one of the possible sources of conflict in the discipline. As early as 1999, the conceptual gap between non-essentialist and empirical approaches came up for debate at the Vic Forum on Training

Translators and Interpreters: New Directions for the Millennium. The discussants, Rosemary Arrojo and Andrew Chesterman, explicitly sought common shared ground for both approaches.

Interdisciplinarity has made the creation of new paradigms possible, as most of the developed theories grew from contact with other disciplines like linguistics, comparative literature, cultural studies, philosophy, sociology or historiography. At the same time, it might have provoked the fragmentation of translation studies as a discipline on its own right.

A second source of conflict rises from the breach between theory and practice. As the prescriptivism of the earlier studies gives room to descriptivism and theorization, professionals see less applicability of the studies. At the same time, university research assessment places little if any importance on translation practice.

Translation studies has shown a tendency to broaden its fields of inquiry, and this trend may be expected to continue. This particularly concerns extensions into adaptation studies, intralingual translation, translation between semiotic systems (image to text to music, for example), and translation as the form of all interpretation and thus of all understanding, as suggested in Roman Jakobson's work, *On Linguistic Aspects of Translation*.

## Chapter 5

# Writing System

A **writing system** is a method of visually representing verbal communication, based on a **script** and a set of rules regulating its use. While both writing and speech are useful in conveying messages, writing differs in also being a reliable form of information storage and transfer. Writing systems require shared understanding between writers and readers of the meaning behind the sets of characters that make up a script. Writing is usually recorded onto a durable medium, such as paper or electronic storage, although non-durable methods may also be used, such as writing on a computer display, on a blackboard, in sand, or by skywriting. Reading a text can be accomplished purely in the mind as an internal process, or expressed orally.

Writing systems can be placed into broad categories such as alphabets, syllabaries, or logographies, although any particular system may have attributes of more than one category. In the alphabetic category, a standard set of letters represent speech sounds. In a syllabary, each symbol correlates to a syllable or mora. In a logography, each character represents a semantic unit such as a word or morpheme. Abjads differ from alphabets in that vowels are not indicated, and in abugidas or alphasyllabaries each character represents a consonant–vowel pairing.

Alphabets typically use a set of less than 100 symbols to fully express a language, whereas syllabaries can have several

hundred, and logographies can have thousands of symbols. Many writing systems also include a special set of symbols known as punctuation which is used to aid interpretation and help capture nuances and variations in the message's meaning that are communicated verbally by cues in timing, tone, accent, inflection or intonation.

Writing systems were preceded by proto-writing, which used pictograms, ideograms and other mnemonic symbols. Proto-writing lacked the ability to capture and express a full range of thoughts and ideas. The invention of writing systems, which dates back to the beginning of the Bronze Age in the late Neolithic Era of the late 4th millennium BC, enabled the accurate durable recording of human history in a manner that was not prone to the same types of error to which oral history is vulnerable. Soon after, writing provided a reliable form of long distance communication. With the advent of publishing, it provided the medium for an early form of mass communication.

# General properties

Writing systems are distinguished from other possible symbolic communication systems in that a writing system is always associated with at least one spoken language. In contrast, visual representations such as drawings, paintings, and non-verbal items on maps, such as contour lines, are not language-related. Some symbols on information signs, such as the symbols for male and female, are also not language related, but can grow to become part of language if they are often used in conjunction with other language elements. Some other symbols, such as numerals and the ampersand, are not

directly linked to any specific language, but are often used in writing and thus must be considered part of writing systems.

Every human community possesses language, which many regard as an innate and defining condition of humanity. However, the development of writing systems, and the process by which they have supplanted traditional oral systems of communication, have been sporadic, uneven and slow. Once established, writing systems generally change more slowly than their spoken counterparts. Thus they often preserve features and expressions which are no longer current in the spoken language. One of the great benefits of writing systems is that they can preserve a permanent record of information expressed in a language.

All writing systems require:

- at least one set of defined base elements or symbols, individually termed *signs* and collectively called a *script*;
- at least one set of rules and conventions (orthography) understood and shared by a community, which assigns meaning to the base elements (graphemes), their ordering and relations to one another;
- at least one language (generally spoken) whose constructions are represented and can be recalled by the interpretation of these elements and rules;
- some physical means of distinctly representing the symbols by application to a permanent or semi-permanent medium, so they may be interpreted

(usually visually, but tactile systems have also been devised).

# Basic terminology

In the examination of individual scripts, the study of writing systems has developed along partially independent lines. Thus, the terminology employed differs somewhat from field to field.

## Text, writing, reading and orthography

The generic term *text* refers to an instance of written or spoken material with the latter having been transcribed in some way. The act of composing and recording a text may be referred to as *writing*, and the act of viewing and interpreting the text as *reading*. *Orthography* refers to the method and rules of observed writing structure (literal meaning, "correct writing"), and particularly for alphabetic systems, includes the concept of *spelling*.

## Grapheme and phoneme

A *grapheme* is a specific base unit of a writing system. They are the *minimally significant* elements which taken together comprise the set of "building blocks" out of which texts made up of one or more writing systems may be constructed, along with rules of correspondence and use. The concept is similar to that of the phoneme used in the study of spoken languages. For example, in the Latin-based writing system of standard contemporary English, examples of graphemes include the

majuscule and minuscule forms of the twenty-six letters of the alphabet (corresponding to various phonemes), marks of punctuation (mostly non-phonemic), and a few other symbols such as those for numerals (logograms for numbers).

An individual grapheme may be represented in a wide variety of ways, where each variation is visually distinct in some regard, but all are interpreted as representing the "same" grapheme. These individual variations are known as *allographs* of a grapheme (compare with the term allophone used in linguistic study). For example, the minuscule letter *a* has different allographs when written as a cursive, block, or typed letter. The choice of a particular allograph may be influenced by the medium used, the writing instrument, the stylistic choice of the writer, the preceding and following graphemes in the text, the time available for writing, the intended audience, and the largely unconscious features of an individual's handwriting.

## Glyph, sign and character

The terms *glyph*, *sign* and *character*are sometimes used to refer to a grapheme. Common usage varies from discipline to discipline; compare cuneiform sign, Maya glyph, Chinese character. The glyphs of most writing systems are made up of lines (or strokes) and are therefore called linear, but there are glyphs in non-linear writing systems made up of other types of marks, such as Cuneiform and Braille.

## Complete and partial writing systems

Writing systems may be regarded as *complete* according to the extent to which they are able to represent all that may be expressed in the spoken language, while a *partial* writing system is limited in what it can convey.

## Writing systems, languages and conceptual systems

Writing systems can be independent from languages, one can have multiple writing systems for a language, e.g., Hindustani; and one can also have one writing system for multiple languages, e.g., the Arabic script. Chinese characters were also borrowed by other countries as their early writing systems, e.g., the early writing systems of Vietnamese language until the beginning of the 20th century.

To represent a conceptual system, one uses one or more languages, e.g., mathematics is a conceptual system and one may use first-order logic and a natural language together in representation.

# History

Writing systems were preceded by proto-writing, systems of ideographic and/or early mnemonic symbols. The best-known examples are:

- "Token system", a recording system used for accounting purposes in Mesopotamia c. 9000 BC
- Jiahu symbols, carved on tortoiseshells in Jiahu, c. 6600 BC
- Vinča symbols (Tărtăria tablets), c. 5300 BC
- Proto-cuneiform c. 3500 BC
- Possibly the early Indus script, c. 3500 BC, as its nature is disputed
- Nsibidi script, c. before 500 AD

The invention of the first writing systems is roughly contemporary with the beginning of the Bronze Age (following the late Neolithic) in the late 4th millennium BC. The Sumerian archaic cuneiform script closely followed by the Egyptian hieroglyphs are generally considered the earliest writing systems, both emerging out of their ancestral proto-literate symbol systems from 3400 to 3200 BC with earliest coherent texts from about 2600 BC. It is generally agreed that the historically earlier Sumerian writing was an independent invention; however, it is debated whether Egyptian writing was developed completely independently of Sumerian, or was a case of cultural diffusion.

A similar debate exists for the Chinese script, which developed around 1200 BC. The Chinese script is probably an independent invention, because there is no evidence of contact between China and the literate civilizations of the Near East, and because of the distinct differences between the Mesopotamian and Chinese approaches to logography and phonetic representation.

The pre-ColumbianMesoamerican writing systems (including among others Olmec and Maya scripts) are generally believed to have had independent origins.

A hieroglyphic writing system used by pre-colonial Mi'kmaq, which was observed by missionaries from the 17th to 19th centuries, is thought to have developed independently. There is some debate over whether or not this was a fully formed system or just a series of mnemonic pictographs.

It is thought that the first consonantal alphabetic writing appeared before 2000 BC, as a representation of language developed by Semitic tribes in the Sinai Peninsula (see History of the alphabet). Most other alphabets in the world today either descended from this one innovation, many via the Phoenician alphabet, or were directly inspired by its design.

The first true alphabet is the Greek script which consistently represents vowels since 800 BC. The Latin alphabet, a direct descendant, is by far the most common writing system in use.

# Functional classification

Several approaches have been taken to classify writing systems, the most common and basic one is a broad division into three categories: *logographic*, *syllabic*, and *alphabetic* (or *segmental*); however, all three may be found in any given writing system in varying proportions, often making it difficult to categorise a system uniquely. The term *complex system* is sometimes used to describe those where the admixture makes classification problematic. Modern linguists regard such approaches, including Diringer's

- pictographic script
- ideographic script
- analytic transitional script
- phonetic script
- alphabetic script

as too simplistic, often considering the categories to be incomparable. Hill split *writing* into three major categories of linguistic analysis, one of which covers discourses and is not usually considered writing proper:

- *discourse system*
- *iconic discourse system*, e.g. Amerindian
- *conventional discourse system*, e.g. Quipu
- *morphemic writing system*, e.g. Egyptian, Sumerian, Maya, Chinese, Anatolian Hieroglyphs
- *phonemic writing system*
- *partial phonemic writing system*, e.g. Egyptian, Hebrew, Arabic
- *poly-phonemic writing system*, e.g. Linear B, Kana, Cherokee
- *mono-phonemic writing system*
- *phonemic writing system*, e.g. Ancient Greek, Old English
- *morpho-phonemic writing system*, e.g. German, Modern English

Sampson draws a distinction between *semasiography* and *glottography*

- semasiography, relating visible marks to meaning directly without reference to any specific spoken language

- glottography, using visible marks to represent forms of a spoken language
- logography, representing a spoken language by assigning distinctive visible marks to linguistic elements of André Martinet's "first articulation" (Martinet 1949), i.e. morphemes or words
- phonography, achieving the same goal by assigning marks to elements of the "second articulation", e.g. phonemes, syllables

DeFrancis, criticizing Sampson's introduction of *semasiographic writing* and *featural alphabets* stresses the phonographic quality of writing proper

- *pictures*
- *nonwriting*
- *writing*
- *rebus*
- *syllabic systems*
- *pure syllabic*, e.g. Linear B, Yi, Kana, Cherokee
- *morpho-syllabic*, e.g. Sumerian, Chinese, Mayan
- *consonantal*
- *morpho-consonantal*, e.g. Egyptian
- *pure consonantal*, e.g. Phoenician
- *alphabetic*
- *pure phonemic*, e.g. Greek
- *morpho-phonemic*, e.g. English

Faber categorizes phonographic writing by two levels, linearity and coding:

- *logographic*, e.g. Chinese, Ancient Egyptian
- *phonographic*

- *syllabically linear*
- *syllabically coded*, e.g. Kana, Akkadian
- *segmentally coded*, e.g. Hebrew, Syriac, Arabic, Ethiopian, Amharic, Devanagari
- *segmentally linear*
- *complete* (alphabet), e.g. Greco-Latin, Cyrillic
- *defective*, e.g. Ugaritic, Phoenician, Aramaic, Old South Arabian, Paleo-Hebrew

## Logographic systems

A *logogram* is a single written character which represents a complete grammatical word. Chinese characters are type examples of logograms.

As each character represents a single word (or, more precisely, a morpheme), many logograms are required to write all the words of language. The vast array of logograms and the memorization of what they mean are considered by some as major disadvantages of logographic systems over alphabetic systems. However, since the meaning is inherent to the symbol, the same logographic system can theoretically be used to represent different languages. In practice, the ability to communicate across languages works best for the closely related varieties of Chinese, and only to a lesser extent for other languages, as differences in syntax reduce the cross linguistic portability of a given logographic system.

Japanese uses Chinese logograms extensively in its writing systems, with most of the symbols carrying the same or similar meanings. However, the grammatical differences between Japanese and Chinese are significant enough that a long

Chinese text is not readily understandable to a Japanese reader without any knowledge of basic Chinese grammar, though short and concise phrases such as those on signs and newspaper headlines are much easier to comprehend. Similarly, a Chinese reader can get a general idea of what a long Japanese text means but usually cannot understand the text fully.

While most languages do not use wholly logographic writing systems, many languages use some logograms. A good example of modern western logograms are the Arabic numerals: everyone who uses those symbols understands what *1* means whether they call it *one*, *eins*, *uno*, *yi*, *ichi*, *ehad*, *ena*, or *jedan*. Other western logograms include the ampersand&, used for *and*, the at sign@, used in many contexts for *at*, the percent sign% and the many signs representing units of currency ($, ¢, •, £, ¥ and so on.)

Logograms are sometimes called ideograms, a word that refers to symbols which graphically represent abstract ideas, but linguists avoid this use, as Chinese characters are often semantic–phonetic compounds, symbols which include an element that represents the meaning and a phonetic complement element that represents the pronunciation. Some nonlinguists distinguish between lexigraphy and ideography, where symbols in lexigraphies represent words and symbols in ideographies represent words or morphemes.

The most important (and, to a degree, the only surviving) modern logographic writing system is the Chinese one, whose characters have been used with varying degrees of modification in varieties of Chinese, Japanese, Korean, Vietnamese, and

other east Asian languages. Ancient Egyptian hieroglyphs and the Mayan writing system are also systems with certain logographic features, although they have marked phonetic features as well and are no longer in current use. Vietnamese switched to the Latin alphabet in the 20th century and the use of Chinese characters in Korean is increasingly rare. The Japanese writing system includes several distinct forms of writing including logography.

## Syllabic systems: syllabary

*Another type of writing system with systematic syllabic linear symbols, the abugidas, is discussed below as well.*

As logographic writing systems use a single symbol for an entire word, a *syllabary* is a set of written symbols that represent (or approximate) syllables, which make up words. A symbol in a syllabary typically represents a consonant sound followed by a vowel sound, or just a vowel alone.

In a "true syllabary", there is no systematic graphic similarity between phonetically related characters (though some do have graphic similarity for the vowels). That is, the characters for /ke/, /ka/ and /ko/ have no similarity to indicate their common "k" sound (voiceless velar plosive). More recent creations such as the Cree syllabary embody a system of varying signs, which can best be seen when arranging the syllabogram set in an onset–coda or onset–rime table.

Syllabaries are best suited to languages with relatively simple syllable structure, such as Japanese. The English language, on the other hand, allows complex syllable structures, with a

relatively large inventory of vowels and complex consonant clusters, making it cumbersome to write English words with a syllabary. To write English using a syllabary, every possible syllable in English would have to have a separate symbol, and whereas the number of possible syllables in Japanese is around 100, in English there are approximately 15,000 to 16,000.

However, syllabaries with much larger inventories do exist. The Yi script, for example, contains 756 different symbols (or 1,164, if symbols with a particular tone diacritic are counted as separate syllables, as in Unicode). The Chinese script, when used to write Middle Chinese and the modern varieties of Chinese, also represents syllables, and includes separate glyphs for nearly all of the many thousands of syllables in Middle Chinese; however, because it primarily represents morphemes and includes different characters to represent homophonous morphemes with different meanings, it is normally considered a logographic script rather than a syllabary.

Other languages that use true syllabaries include MycenaeanGreek (Linear B) and Indigenous languages of the Americas such as Cherokee. Several languages of the Ancient Near East used forms of cuneiform, which is a syllabary with some non-syllabic elements.

## Segmental systems: alphabets

An *alphabet* is a small set of *letters* (basic written symbols), each of which roughly represents or represented historically a segmentalphoneme of a spoken language. The word *alphabet*is

derived from alpha and beta, the first two symbols of the Greek alphabet.

The first type of alphabet that was developed was the abjad. An abjad is an alphabetic writing system where there is one symbol per consonant. Abjads differ from other alphabets in that they have characters only for consonantal sounds. Vowels are not usually marked in abjads. All known abjads (except maybe Tifinagh) belong to the Semitic family of scripts, and derive from the original Northern Linear Abjad. The reason for this is that Semitic languages and the related Berber languages have a morphemic structure which makes the denotation of vowels redundant in most cases.

Some abjads, like Arabic and Hebrew, have markings for vowels as well. However, they use them only in special contexts, such as for teaching. Many scripts derived from abjads have been extended with vowel symbols to become full alphabets. Of these, the most famous example is the derivation of the Greek alphabet from the Phoenician abjad. This has mostly happened when the script was adapted to a non-Semitic language. The term *abjad* takes its name from the old order of the Arabic alphabet's consonants 'alif, bā', jīm, dāl, though the word may have earlier roots in Phoenician or Ugaritic. "Abjad" is still the word for alphabet in Arabic, Malay and Indonesian.

An abugida is an alphabetic writing system whose basic signs denote consonants with an inherent vowel and where consistent modifications of the basic sign indicate other following vowels than the inherent one. Thus, in an abugida there may or may not be a sign for "k" with no vowel, but also one for "ka" (if "a" is the inherent vowel), and "ke" is written by

modifying the "ka" sign in a way that is consistent with how one would modify "la" to get "le". In many abugidas the modification is the addition of a vowel sign, but other possibilities are imaginable (and used), such as rotation of the basic sign, addition of diacritical marks and so on.

The contrast with "true syllabaries" is that the latter have one distinct symbol per possible syllable, and the signs for each syllable have no systematic graphic similarity. The graphic similarity of most abugidas comes from the fact that they are derived from abjads, and the consonants make up the symbols with the inherent vowel and the new vowel symbols are markings added on to the base symbol. In the Ge'ez script, for which the linguistic term *abugida*was named, the vowel modifications do not always appear systematic, although they originally were more so.

Canadian Aboriginal syllabicscan be considered abugidas, although they are rarely thought of in those terms. The largest single group of abugidas is the Brahmic family of scripts, however, which includes nearly all the scripts used in India and Southeast Asia. The name *abugida*is derived from the first four characters of an order of the Ge'ez script used in some contexts. It was borrowed from Ethiopian languages as a linguistic term by Peter T. Daniels.

## Featural systems

A *featural* script represents finer detail than an alphabet. Here symbols do not represent whole phonemes, but rather the elements (features) that make up the phonemes, such as voicing or its place of articulation. Theoretically, each feature

could be written with a separate letter; and abjads or abugidas, or indeed syllabaries, could be featural, but the only prominent system of this sort is Koreanhangul. In hangul, the featural symbols are combined into alphabetic letters, and these letters are in turn joined into syllabic blocks, so that the system combines three levels of phonological representation.

Many scholars, e.g. John DeFrancis, reject this class or at least labeling hangul as such. The Korean script is a conscious script creation by literate experts, which Daniels calls a "sophisticated grammatogeny". These include stenographies and constructed scripts of hobbyists and fiction writers (such as Tengwar), many of which feature advanced graphic designs corresponding to phonologic properties. The basic unit of writing in these systems can map to anything from phonemes to words. It has been shown that even the Latin script has sub-character "features".

## Ambiguous systems

Most writing systems are not purely one type. The English writing system, for example, includes numerals and other logograms such as #, $, and &, and the written language often does not match well with the spoken one. As mentioned above, all logographic systems have phonetic components as well, whether along the lines of a syllabary, such as Chinese ("logo-syllabic"), or an abjad, as in Egyptian ("logo-consonantal").

Some scripts, however, are truly ambiguous. The semi-syllabaries of ancient Spain were syllabic for plosives such as *p*, *t*, *k*, but alphabetic for other consonants. In some versions, vowels were written redundantly after syllabic letters,

conforming to an alphabetic orthography. Old Persian cuneiform was similar. Of 23 consonants (including null), seven were fully syllabic, thirteen were purely alphabetic, and for the other three, there was one letter for /C*u*/ and another for both /C*a*/ and /C*i*/. However, all vowels were written overtly regardless; as in the Brahmic abugidas, the /C*a*/ letter was used for a bare consonant.

The zhuyin phonetic glossing script for Chinese divides syllables in two or three, but into onset, medial, and rime rather than consonant and vowel. Pahawh Hmong is similar, but can be considered to divide syllables into either onset-rime or consonant-vowel (all consonant clusters and diphthongs are written with single letters); as the latter, it is equivalent to an abugida but with the roles of consonant and vowel reversed. Other scripts are intermediate between the categories of alphabet, abjad and abugida, so there may be disagreement on how they should be classified.

# Graphic classification

Perhaps the primary graphic distinction made in classifications is that of *linearity*. Linear writing systems are those in which the characters are composed of lines, such as the Latin alphabet and Chinese characters. Chinese characters are considered linear whether they are written with a ball-point pen or a calligraphic brush, or cast in bronze. Similarly, Egyptian hieroglyphs and Maya glyphs were often painted in linear outline form, but in formal contexts they were carved in bas-relief. The earliest examples of writing are linear: the Sumerian script of c. 3300 BC was linear, though its cuneiform

descendants were not. Non-linear systems, on the other hand, such as braille, are not composed of lines, no matter what instrument is used to write them.

Cuneiform was probably the earliest non-linear writing. Its glyphs were formed by pressing the end of a reed stylus into moist clay, not by tracing lines in the clay with the stylus as had been done previously. The result was a radical transformation of the appearance of the script.

Braille is a non-linear adaptation of the Latin alphabet that completely abandoned the Latin forms. The letters are composed of raised bumps on the writing substrate, which can be leather (Louis Braille's original material), stiff paper, plastic or metal.

There are also transient non-linear adaptations of the Latin alphabet, including Morse code, the manual alphabets of various sign languages, and semaphore, in which flags or barsare positioned at prescribed angles. However, if "writing" is defined as a potentially permanent means of recording information, then these systems do not qualify as writing at all, since the symbols disappear as soon as they are used. (Instead, these transient systems serve as signals.)

# Directionality

Scripts are graphically characterized by the direction in which they are written. Egyptian hieroglyphs were written either left to right or right to left, with the animal and human glyphs turned to face the beginning of the line. The early alphabet could be written in multiple directions: horizontally (side to

side), or vertically (up or down). Prior to standardization, alphabetical writing was done both left-to-right (LTR or sinistrodextrally) and right-to-left (RTL or dextrosinistrally). It was most commonly written boustrophedonically: starting in one (horizontal) direction, then turning at the end of the line and reversing direction.

The Greek alphabet and its successors settled on a left-to-right pattern, from the top to the bottom of the page. Other scripts, such as Arabic and Hebrew, came to be writtenright-to-left. Scripts that incorporate Chinese characters have traditionally been written vertically (top-to-bottom), from the right to the left of the page, but nowadays are frequently written left-to-right, top-to-bottom, due to Western influence, a growing need to accommodate terms in the Latin script, and technical limitations in popular electronic document formats.

Chinese characters sometimes, as in signage, especially when signifying something old or traditional, may also be written from right to left. The Old Uyghur alphabet and its descendants are unique in being written top-to-bottom, left-to-right; this direction originated from an ancestral Semitic direction by rotating the page 90° counter-clockwise to conform to the appearance of vertical Chinese writing.

Several scripts used in the Philippines and Indonesia, such as Hanunó'o, are traditionally written with lines moving away from the writer, from bottom to top, but are read horizontally left to right; however, Kulitan, another Philippine script, is written top to bottom and right to left. Oghamis written bottom to top and read vertically, commonly on the corner of a stone.

Left-to-right writing has the advantage that since most people are right-handed, the hand does not interfere with the just-written text, which might not yet have dried, since the hand is on the right side of the pen.

# On computers

In computers and telecommunication systems, writing systems are generally not codified as such, but graphemes and other grapheme-like units that are required for text processing are represented by "characters" that typically manifest in encoded form. There are many character encoding standards and related technologies, such as ISO/IEC 8859-1 (a character repertoire and encoding scheme oriented toward the Latin script), CJK (Chinese, Japanese, Korean) and bi-directional text.

Today, many such standards are re-defined in a collective standard, the ISO/IEC 10646 "Universal Character Set", and a parallel, closely related expanded work, *The Unicode Standard*. Both are generally encompassed by the term Unicode. In Unicode, each character, in every language's writing system, is (simplifying slightly) given a unique identification number, known as its *code point*. Computer operating systems use code points to look up characters in the font file, so the characters can be displayed on the page or screen.

A keyboard is the device most commonly used for writing via computer. Each key is associated with a standard code which the keyboard sends to the computer when it is pressed. By using a combination of alphabetic keys with modifier keys such

as Ctrl, Alt, Shift and AltGr, various character codes are generated and sent to the CPU. The operating system intercepts and converts those signals to the appropriate characters based on the keyboard layout and input method, and then delivers those converted codes and characters to the running application software, which in turn looks up the appropriate glyph in the currently used font file, and requests the operating system to draw these on the screen.

Chapter 6

# Theoretical Frameworks

## Phrase structure grammar

The term **phrase structure grammar** was originally introduced by Noam Chomsky as the term for grammar studied previously by Emil Post and Axel Thue (Post canonical systems). Some authors, however, reserve the term for more restricted grammars in the Chomsky hierarchy: context-sensitive grammars or context-free grammars. In a broader sense, phrase structure grammars are also known as *constituency grammars*. The defining trait of phrase structure grammars is thus their adherence to the constituency relation, as opposed to the dependency relation of dependency grammars.

## Constituency relation

In linguistics, phrase structure grammars are all those grammars that are based on the constituency relation, as opposed to the dependency relation associated with dependency grammars; hence, phrase structure grammars are also known as constituency grammars. Any of several related theories for the parsing of natural language qualify as constituency grammars, and most of them have been developed from Chomsky's work, including

- Government and binding theory

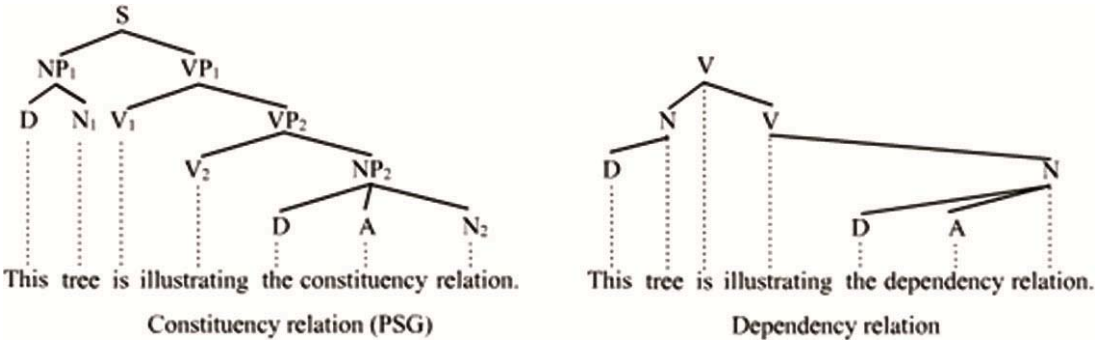
- Generalized phrase structure grammar
- Head-driven phrase structure grammar
- Lexical functional grammar
- The minimalist program
- Nanosyntax

Further grammar frameworks and formalisms also qualify as constituency-based, although they may not think of themselves as having spawned from Chomsky's work, e.g.

- Arc pair grammar, and
- Categorial grammar.

The fundamental trait that these frameworks all share is that they view sentence structure in terms of the constituency relation. The constituency relation derives from the subject-predicate division of Latin and Greek grammars that is based on term logic and reaches back to Aristotle in antiquity. Basic clause structure is understood in terms of a binary division of the clause into subject (noun phrase NP) and predicate (verb phrase VP).

The binary division of the clause results in a one-to-one-or-more correspondence. For each element in a sentence, there are one or more nodes in the tree structure that one assumes for that sentence. A two word sentence such as *Luke laughed* necessarily implies three (or more) nodes in the syntactic structure: one for the noun *Luke* (subject NP), one for the verb *laughed* (predicate VP), and one for the entirety *Luke laughed* (sentence S). The constituency grammars listed above all view sentence structure in terms of this one-to-one-or-more correspondence.




- Constituency relation (PSG) Dependency relation

# Dependency relation

By the time of GottlobFrege, a competing understanding of the logic of sentences had arisen. Frege rejected the binary division of the sentence and replaced it with an understanding of sentence logic in terms of logical predicates and their arguments. On this alternative conception of sentence logic, the binary division of the clause into subject and predicate was not possible. It therefore opened the door to the dependency relation (although the dependency relation had also existed in a less obvious form in traditional grammars long before Frege). The dependency relation was first acknowledged concretely and developed as the basis for a comprehensive theory of syntax and grammar by Lucien Tesnière in his posthumously published work *Éléments de syntaxestructurale* (Elements of Structural Syntax).

The dependency relation is a one-to-one correspondence: for every element (word or morph) in a sentence, there is just one node in the syntactic structure. The distinction is thus a graph-theoretical distinction. The dependency relation restricts the number of nodes in the syntactic structure of a sentence to the exact number of syntactic units (usually words) that that sentence contains. Thus the two word sentence *Luke laughed*

implies just two syntactic nodes, one for *Luke* and one for *laughed*. Some prominent dependency grammars are listed here:

- Recursive categorical syntax, sometimes called *algebraic syntax*
- Functional generative description
- Lexicase
- Link grammar
- Meaning-text theory
- Operator grammar
- Word grammar

Since these grammars are all based on the dependency relation, they are by definition NOT phrase structure grammars.

# Non-descript grammars

Other grammars generally avoid attempts to group syntactic units into clusters in a manner that would allow classification in terms of the constituency vs. dependency distinction. In this respect, the following grammar frameworks do not come down solidly on either side of the dividing line:

- Construction grammar
- Cognitive grammar

# Distributionalism

**Distributionalism** was a general theory of language and a discovery procedure for establishing elements and structures of language based on observed usage. It can be seen as an elaboration of structuralism but takes a more computational approach. Originally mostly applied to understanding phonological processes and phonotactics, distributional methods were also applied to work on lexical semantics and provide the basis for the distributional hypothesis for meaning. Current computational approaches to learn the semantics of words from text in the form of word embeddings using machine learningare based on distributional theory.

# Origins

Distributionalismcan be said to have originated in the work of structuralist linguist Leonard Bloomfield and was more clearly formalised by Zellig S. Harris. This theory emerged in the United States in the 1950s, as a variant of structuralism, which was the mainstream linguistic theory at the time, and dominated American linguistics for some time. Using "distribution" as a technical term for a component of discovery procedure is likely first to have been done by Morris Swadesh in 1934 and then applied to principles of phonematics, to establish which observable various sounds of a language constitute the allophones of a phoneme and which should be kept as separate phonemes. According to Turenne and Pomerol, distributionalism was in fact a second phase in the history of linguistics, following that of structuralism, as

distributionalism was mainly dominant since 1935 to 1960. It is considered one of the scientific grounds of Noam Chomsky's generative grammar and had considerable influence on language teaching.

Distributionalism has much in common with structuralism. However, both appear in the United States while the theses of Ferdinand de Saussure are only just beginning to be known in Europe: distributionism must be considered as an original theory in relation to Saussurianism.

Behaviorist psychological theories which allowed the birth of distributionalism are reminiscent of Pavlov's work on animals. According to these theories, human behaviour would be totally explainable, and its mechanics could be studied. The study of reflexes, for example, should have made it possible to predict certain attitudes. Leonard Bloomfield argues that language, like behaviour, could be analysed as a predictable mechanism, explicable by the external conditions of its appearance.

The notions of "mechanism", "inductive method" and "corpus" are key terms of distributionalism.

## Mechanism vs Mentalism

Bloomfield calls his thesis *mechanism*, and he opposes it to mentalism: for him, in fact, speech cannot be explained as an effect of thoughts (intentions, beliefs, feelings). Thus, one must be able to account for linguistic behaviour and the hierarchical structure of the messages conveyed without any assumptions about the speakers' intentions and mental states.

From the behaviourist perspective, a given stimulus corresponds to a given response. However, meaning is an unstable thing for distributionists, depending on the situation, and is not observable. It must therefore be eliminated as an element of language analysis. The only regularity is of a morphosyntactic nature: it is the structural invariants of the morphosyntax that allow us to reconstruct the language system from an analysis of its observable elements, the words of a given corpus.

# Salient features

The main idea of distributionalism is that linguistic units "are what they do", which means that the identity of linguistic units are *defined by their distribution*. Zellig Harris used to consider meaning as too intuitive to be a reliable ground for linguistic research. Language use has to be observeddirectly while looking at all the environments in which a unit can occur. Harris advocated for a distributional approach, since "difference of meaning correlates with difference of distribution.".

# Prague linguistic circle

The **Prague school** or **Prague linguistic circle** is a language and literature society. It started in 1926 as a group of linguists, philologists and literary critics in Prague. Its proponents developed methods of structuralist literary analysis and a theory of the standard language and of language cultivation from 1928 to 1939. The linguistic circle was

founded in the Café Derby in Prague, which is also where meetings took place during its first years.

The Prague School has had a significant continuing influence on linguistics and semiotics. After the Czechoslovak coup d'état of 1948, the circle was disbanded in 1952, but the Prague School continued as a major force in linguistic functionalism (distinct from the Copenhagen school or English Firthian – later Hallidean – linguistics). The American scholar Dell Hymes cites his 1962 paper, "The Ethnography of Speaking," as the formal introduction of Prague functionalism to American linguistic anthropology. The Prague structuralists also had a significant influence on structuralist film theory, especially through the introduction of the ostensive sign.

Today the Prague linguistic circle is a scholarly society which aims to contribute to the knowledge of language and related sign systems according to functionally structural principles. To this end, it organizes regular meetings with lectures and debates, publishes professional publications, and organizes international meetings.

# History

The Prague linguistic circle included the Russian émigrés Roman Jakobson, Nikolai Trubetzkoy, and Sergei Karcevskiy, as well as the famous Czech literary scholars René Wellek and Jan Mukařovský. The instigator of the circle, and its first president until his death in 1945, was the Czech linguist VilémMathesius.

In 1929 the Circle promulgated its theses in a paper submitted to the First Congress of Slavists. "The programmatic 1929 Prague *Theses*, surely one of the most imposing linguistic edifices of the 20th century, incapsulated [sic] the functionalist credo." In the late 20th century, English translations of the Circle's seminal works were published by the Czech linguist Josef Vachek in several collections.

Also in 1929, the group launched a journal, *Travaux du CercleLinguistique de Prague*. World War II brought an end to it. The *Travaux*was briefly resurrected in 1966–1971. The inaugural issue was devoted to the political science concept of center and periphery. It was resurrected yet again in 1995. The group's Czech language work is published in *Slovo a slovesnost* (Word and Literature).

# Members

- Petr Bogatyrev [cs; ru]
- FrantišekČermák [cs; pl]
- Miroslav Červenka [cs; pl; de]
- BohuslavHavránek
- TomášHoskovec [cs; pl]
- Josef Hrabák [cs; pl; ru]
- Roman Jakobson
- SergejKarcevskij [cs; ru]
- OldřichLeška [cs]
- Alena Macurová [cs]
- VilémMathesius
- Jan Mukařovský
- Karel Oliva [cs]

- VladimírSkalička
- BohumilTrnka [cs]
- Pavel Trost [cs; de; pl]
- Nikolai Trubetzkoy
- Josef Vachek [cs; pl]
- JiříVeltruský
- MilošWeingart [cs; pl]
- René Wellek
- Ludwig Winder
- Contributors
- AleksandarBelić, president of the Serbian Academy of Sciences and Arts
- ÉmileBenveniste
- Karl Bühler
- Albert Willem de Groot [ru]
- Daniel Jones
- André Martinet
- LadislavMatejka
- Lucien Tesnière
- Valentin Voloshinov
- Influences
- Jan Baudouin de Courtenay
- FilippFortunatov [ru], the founder of the Moscow linguistic circle
- Anton Marty
- Ferdinand de Saussure
- Influenced
- Noam Chomsky
- Joseph Greenberg
- JiříLevý
- Dell Hymes
- Alf Sommerfelt

- JožeToporišič
- Michael Halliday
- Viktor Shklovsky
- Emilio AlarcosLlorach [es]
- Michael Silverstein
- Jan Firbas
- LubomírDoležel
- Austin Warren
- Jan Baudouin de Courtenay
- Louis Hjelmslev
- Jaroslav Vacek
- Jaroslav Peregrin
- Miroslav Komárek
- Tartu–Moscow Semiotic School

# Cognitive grammar

**Cognitive grammar** is a cognitive approach to language developed by Ronald Langacker, which hypothesizes that grammar, semantics, and lexicon exist on a continuum instead of as separate processes altogether. This approach to language was one of the first projects of cognitive linguistics. In this system, grammar is not a formal system operating independently of meaning. Rather, grammar is itself meaningful and inextricable from semantics.

Construction grammar is a similar foci of cognitive approaches to grammar. While cognitive grammar emphasizes the study of the cognitive principles that give rise to linguistic organization, construction grammar aims to provide a more descriptively and

formally detailed account of the linguistic units that comprise a particular language.

Langacker first explicates the system of cognitive grammar in his seminal, two-volume work *Foundations of Cognitive Grammar*. Volume one is titled "Theoretical Prerequisites", and it explores Langacker's hypothesis that grammar may be deconstructed into patterns that come together in order to represent concepts. This volume concentrates on the broad scope of language especially in terms of the relationship between grammar and semantics. Volume two is titled "Descriptive Application", as it moves beyond the first volume to elaborate on the ways in which Langacker's previously described theories may be applied. Langacker invites his reader to utilize the tools presented in *Foundations'* first volume in a wide range of, mainly English, grammatical situations.

# Theory

Cognitive grammar is unorthodox with respect to generative grammars and American structuralism. It primarily diverges from Chomskyan tradition through its assertion that grammar and language are integral and essential parts of cognition, not merely autonomous processes in the brain. Langacker argues not only that cognitive grammar is natural by virtue of its psychological plausibility, but also that it offers conceptual unification and theoretical austerity. It considers the basic units of language to be symbols (i.e. conventional pairings of a semantic structure with a phonological label). Grammar consists of constraints on how these units can be combined to

generate larger phrases. The semantic aspects of cognitive grammar are modeled as image schemas rather than propositions, although these schema are only demonstrative, and are not intended to reflect any actual visual operation occurring during the production and perception of language. A consequence of the interrelation between semantic structure and phonological label is that each can invoke the other.

# Usage-based models of language

The **Usage-based linguistics** is a linguistics approach within a broader functional/cognitive framework, that emerged since the late 1980s, and that assumes a profound relation between linguistic structure and usage. It challenges the dominant focus, in 20th century linguistics (and in particular con formalism-generativism), on considering language as an isolated system removed from its use in human interaction and human cognition. Rather, usage-based models posit that linguistic information is expressed via context-sensitive mental processing and mental representations, which have the cognitive ability to succinctly account for the complexity of actual language use at all levels (phonetics and phonology, morphology and syntax, pragmatics and semantics). Broadly speaking, a usage-based model of language accounts for language acquisition and processing, synchronic and diachronic patterns, and both low-level and high-level structure in language, by looking at actual language use.

The term *usage-based* was coined by Ronald Langacker in 1987. Usage-based models of language have become a significant new trend in linguistics since the early 2000s.

Influential proponents of usage-based linguistics include Michael Tomasello, Joan Bybee and Morten Christiansen.

Together with related approaches, such as construction grammar, emergent grammar, and language as a complex adaptive system, usage-based linguistics belongs to the wider framework of evolutionary linguistics. It studies the lifespan of linguistic units (e.g. words, suffixes), arguing that they can survive language change through frequent usage or by participating in usage-based generalizations if their syntactic, semantic or pragmatic features overlap with other similar constructions. There is disagreement whether the approach is different from memetics or essentially the same.

# Disciplinary roots

### West coast cognitive functionalism

West Coast cognitive functionalism (WCCF) played a major role in the creation of the usage-based enterprise. Firstly, a crucial point in WCCF was Eleanor Rosch's paper on semantic categories in human cognition, which studied fuzzy semantic categories with central and peripheral concepts. Subsequently, Robin Lakoff (1987) applied these concepts to linguistic studies. For usage-based models of language, these discoveries legitimized interest in the peripheral phenomena and inspired the examination of the ontological status of the rules themselves. Secondly, WCCF focuses on the effects of social/ textual context and cognitive processes on human thought, instead of established systems and representations, which motivated the study of external sources in usage-based

language research. For example, in analyzing the differences between the grammatical notions of subject vs. topic, Li and Thompson (1976), found that the repetition of certain topics by a speech community resulted in the surfacing and crystallization of formal properties into syntactic entities, namely the subject. This notion of syntax and morphology being an outcome of pragmatic and cognitive factors was influential in the development of usage-based models. Thirdly, the WCCF methodology of linguistic typologyis similarly practised in usage-based models, in collecting data from real communicative contexts and analyzing them for typological regularities. This highlights an important aspect of usage-based research, the study of methods for the integration of synchrony and diachrony.

**Langacker's Cognitive Grammar**

The term 'usage-based' was coined by Ronald Langacker in 1987, while doing research on Cognitive Grammar. Langacker identified commonly recurring linguistic patterns (patterns such as those associated with Wh- fronting, subject-verb agreement, the use of present participles, etc.) and represented these supposed rule-governed behaviours on a hierarchical structure. The Cognitive Grammar model represented grammar, semantics and lexicon as associated processes that were laid on a continuum, which provided a theoretical framework that was significant in studying the usage-based conception of language. Consequently, a usage-based model accounts for these rule-governed language behaviours by providing a representational scheme that is entirely instance-based, and able to recognize and uniquely represent each familiar pattern, which occurs with varying strengths at different instances. His

usage-based model draws on the cognitive psychology of schemata, which are flexible hierarchical structures that are able to accommodate the complexity of mental stimuli. Similarly, as humans perceive linguistic abstractions as multilayered, ranging from patterns that occur across whole utterances to those that occur in phonetic material, the usage-based model acknowledges the differing levels of granularity in speakers' knowledge of their language. Langacker's work emphasizes that both abstract structure and instance-based detail are contained in language, differing in granularity but not in basic principles.

**Bybee's Dynamic Usage-based framework**

Bybee's work greatly inspired the creation of usage-based models of language. Bybee's model makes predictions about and explains synchronic, diachronic and typological patterns within languages, such as which variants will occur in which contexts, what forms they will take, and about their diachronic consequences. Using the linguistic phenomenon of splits (when a word starts to show subtle polysemy, and morphological possibilities for the originally single form ensue), Bybee proves that even irreducibly irregular word-forms are seen to be non-arbitrary when the context it occurs in is taken into consideration in the very representation of morphology. Simultaneously, she shows that even seemingly regular allomorphy is context-sensitive. Splits also aligns with the idea that linguistic forms cannot be studied as isolated entities, but rather in relation to the strength of their attachment to other entities.

## Schmid's Entrenchment-and-Conventionalization model

Hans-JörgSchmid's "Entrenchment-and-Conventionalization" Model offers a comprehensive recent summary approach to usage-based thinking. In great detail and with reference to many sub-disciplines and concepts in linguistics he shows how usage mediates between entrenchment, the establishment of linguistic habits in individuals via repetition and associations, and conventionalization, a continuous feedback cycle which builds shared collective linguistic knowledge. All three components connect linguistic utterance types with their respective situative settings and extralinguistic associations.

# Constructions: Form-meaning pairings

Constructions have direct pairing of form to meaning without intermediate structures, making them appropriate for usage-based models. The usage-based model adopts constructions as the basic unit of form-meaning correspondence. A construction is commonly regarded to be a conventionalized string of words. A key feature of a grammar based on constructions is that it can reflect the deeply intertwined lexical items and grammar structure.

From a grammarian perspective, constructions are groupings of words with idiosyncratic behaviour to a certain extent. They

mostly take on an unpredictable meaning or pragmatic effect, or are formally special. From a broader perspective, construction can also be seen as processing units or chunks, such as sequences of words (or morphemes) which have been used often enough to be accessed together. This implicates that common words sequences are sometimes constructions even if they do not have idiosyncrasies or form. Additionally, chunks or conventionalized sequences can tend to develop special pragmatic implications that can lead to special meaning over time. They can also develop idiosyncrasies of form in a variety of ways.

- It drives me crazy.
- The death of his wife the following year drove him mad.
- This room drives me up the wall.

Adjectives shown here include crazy, mad, and up the wall, which are semantically related to the word drive. In exemplar models, the idea that memory for linguistic experience is similar to memory for other types of memories is proposed. Every token of linguistic experience impacts cognitive representation. And when stored representations are accessed, the representations change. Additionally, memory storage can store detailed information about processed tokens during linguistic experience, including form and context that these tokens were used. In this model, general categories and grammar units can emerge from linguistic experiences stored in memories, as exemplars are categorized by similarity to each other. Contiguous experiences such as meaning and acoustic shape are also recorded to be linked to each other.

**Constructions as Chunks**

By these means repeated sequences become more fluent. Within a chunk, sequential links are graded in strength based on the frequency of the chunk or perhaps the transitions between the elements of a chunk. A construction is a chunk even though it may contain schematic slots, that is, the elements of a chunk can be interrupted.

Memory storage requires links to connect idiomatic phrases together. In chunking, repeated sequences are represented together as units which can be accessed directly. Through this, repeated sequences are more frequent. Sequential links are assessed in strength based on the frequency of the chunk or transitions between elements within a chunk. Additionally, the individual elements of a chunk can link to elements in other contexts. The example of 'drive someone crazy' forms a chunk, however items that compose it are not analyzable individually as words that occur elsewhere in cognitive representation. As chunks are used more frequently, words can lose their associations with exemplars of the same word. This is known as de-categorialization.

# Structural linguistics

**Structural linguistics,** or **structuralism**, in linguistics, denotes schools or theories in which language is conceived as a self-contained, self-regulating semiotic system whose elements are defined by their relationship to other elements within the system. It is derived from the work of Swiss linguist Ferdinand de Saussure and is part of the overall approach of

structuralism. Saussure's *Course in General Linguistics*, published posthumously in 1916, stressed examining language as a dynamic system of interconnected units. Saussure is also known for introducing several basic dimensions of semiotic analysis that are still important today. Two of these are his key methods of syntagmatic and paradigmatic analysis, which define units syntactically and lexically, respectively, according to their contrast with the other units in the system.

*Structuralism* as a term, however, was not used by Saussure who himself called the approach *semiology*. The term *structuralism* is derived from SociologistÉmile Durkheim's anti-Darwinian modification of Herbert Spencer's organic analogy which draws a parallel between social structures and the organs of an organism which have different functions or purposes. Similar analogies and metaphorswere used in the historical-comparative linguistics that Saussure was part of. Saussure himself made a modification of August Schleicher's language–species analogy, based on William Dwight Whitney's critical writings, to turn focus to the internal elements of the language organism, or system. Nonetheless, structural linguistics became mainly associated with Saussure's notion of language as a dual interactive system of symbols and concepts. The term structuralism was adopted to linguistics after Saussure's death by the Prague school linguists Roman Jakobson and Nikolai Trubetzkoy; while the term structural linguistics was coined by Louis Hjelmslev.

# History

Structural linguistics begins with the posthumous publication of Ferdinand de Saussure's *Course in General Linguistics* in 1916, which his students compiled from his lectures. The book proved to be highly influential, providing the foundation for both modern linguistics and semiotics. Structuralist linguistics is often thought of as giving rise to independent European and American traditions due to ambiguity in the term. It is most commonly thought that structural linguistics stems from Saussure's writings; but these were rejected by an American school of linguistics based on Wilhelm Wundt's structural psychology.

## European structuralism

In Europe, Saussure influenced: (1) the Geneva School of Albert Sechehaye and Charles Bally, (2) the Prague linguistic circle, (3) the Copenhagen School of Louis Hjelmslev, (4) the Paris School of André Martinet and Algirdas Julien Greimas, and the Dutch school of Simon Dik. Structural linguistics also had an influence on other disciplines of humanities bringing about the movement known as structuralism.

## 'American structuralism', or American descriptivism

Some confusion is caused by the fact that an American school of linguistics of 1910s through 1950s, which was based on

structural psychology, (especially Wilhelm Wundt's Völkerpsychologie); and later on behavioural psychology, is sometimes nicknamed 'American structuralism'. This framework was not structuralist in the Saussurean sense that it did not consider language as arising from the interaction of meaning and expression. Instead, it was thought that the civilised human mind is organised into binary branching structures. Advocates of this type of structuralism are identified from their use of 'philosophical grammar' with its convention of placing the object, but not the subject, into the verb phrase; whereby the structure is disconnected from semantics in sharp contrast to Saussurean structuralism. This American school is alternatively calleddistributionalism, 'American descriptivism', or the 'Bloomfieldian' school – or 'post-Bloomfieldian', following the death of its leader Leonard Bloomfield in 1949. Nevertheless, Wundt's ideas had already been imported from Germany to American humanities by Franz Boas before him, influencing linguists such as Edward Sapir.

Bloomfield named his psychological approach *descriptive* or *philosophical–descriptive;* as opposed to the historical–comparative study of languages. Structural linguists like Hjelmslev considered his work fragmentary because it eluded a full account of language. The concept of autonomy is also different: while structural linguists considered semiology (the bilateral sign system) as an autonomous system, American descriptivists argued for the autonomy of syntax from semantics. All in all, there were unsolvable incompatibilities between the psychological and positivistic orientation of the Bloomfieldian school, and the semiotic orientation of the structuralists proper. In the generative or Chomskyan concept, a purported rejection of 'structuralism' usually refers to Noam

Chomsky's opposition to the *behaviourism* of Bloomfield's 1933 textbook *Language*; though, coincidentally, he is also opposed to structuralism proper.

# Basic theories and methods

The foundation of structural linguistics is a *sign*, which in turn has two components: a "signified" is an idea or concept, while the "signifier" is a means of expressing the signified. The "sign", e.g. a word, is thus the combined association of signifier and signified. The value of a sign can be defined only by being placed in contrast with other signs. This forms the basis of what later became the paradigmatic dimension of semiotic organization (i.e., terms and inventories of terms that stand in opposition to each other). This is contrasted drastically with the idea that linguistic structures can be examined in isolation from meaning, or that the organisation of the conceptual system can exist without a corresponding organisation of the signifying system.

Paradigmatic relations hold among sets of units, such as the set distinguished phonologically by variation in their initial sound *cat, bat, hat, mat, fat*, or the morphologically distinguished set *ran, run, running*. The units of a set must have something in common with one another, but they must contrast too, otherwise they could not be distinguished from each other and would collapse into a single unit, which could not constitute a set on its own, since a set always consists of more than one unit. Syntagmatic relations, in contrast, are concerned with how units, once selected from their

paradigmatic sets of oppositions, are 'chained' together into structural wholes.

Syntagmatic and paradigmatic relations provide the structural linguist with a tool for categorization for phonology, morphology and syntax. Take morphology, for example. The signs *cat* and *cats* are associated in the mind, producing an abstract paradigm of the word forms of *cat*. Comparing this with other paradigms of word forms, we can note that, in English, the plural often consists of little more than adding an -s to the end of the word. Likewise, through paradigmatic and syntagmatic analysis, we can discover the syntax of sentences. For instance, contrasting the syntagma*je dois* ("I should") and *dois je?* ("Should I?") allows us to realize that in French we only have to invert the units to turn a statement into a question. We thus take syntagmatic evidence (difference in structural configurations) as indicators of paradigmatic relations (e.g., in the present case: questions vs. assertions).

The most detailed account of the relationship between a paradigmatic organisation of language as a motivator and classifier for syntagmatic configurations was provided by Louis Hjelmslev in his *Prolegomena to a Theory of Language*, giving rise to formal linguistics. Hjelmslev's model was subsequently incorporated into systemic functional grammar, functional discourse grammar, and Danish functional grammar.

# Structural explanation

The structural approach in humanities follows from 19th century Geistthinking which is derived from Georg Wilhelm

Friedrich Hegel's philosophy. According to such theories, society or language arises as the collective psyche of a community; and this psyche is sometimes described as an 'organism'. In sociology, Émile Durkheim made a humanistic modification of Herbert Spencer's organic analogy. Durkheim, following Spencer's theory, compared society to an organism which has structures (organs) that carry out different functions. For Durkheim a structural explanation of society is that the population growth, through an organic solidarity (unlike Spencer who believes it happens by a self-interested conduct) leads to an increase of complexity and diversity in a community, creating a society. The structuralist reference became essential when linguistic 'structuralism' was established by the Prague linguistic circle after Saussure's death, following a shift from structural to functional explanation in the social anthropology of Alfred Radcliffe-Brown and Bronisław Malinowski.

Saussure himself had actually used a modification of August Schleicher's Darwinian organic analogy in linguistics; his concept of *la langue* is the social organism or spirit. It needs to be noted that, despite certain similarities, structuralism and functionalism in humanistic linguistics are explicitly anti-Darwinian. This means that linguistic structures are not explained in terms of selection through competition; and that the biological metaphor is not to be taken literally. What is more, Saussure abandoned evolutionary linguistics altogether and, instead, defined synchronic analysis as the study of the language system; and diachronic analysis as the study of language change. With such precaution, structural explanation of language is analogous to structuralism in biology which explains structures in relation with material factors or

substance. In Saussure's explanation, structure follows from systemic consequences of the association of meaning and expression. This can be contrasted with functional explanation which explains linguistic structure in relation to the "adaptation" of language to the community's communicative needs.

Hjelmslev's elaboration of Saussure's structural explanation is that language arises from the structuring of content and expression. He argues that the nature of language could only be understood via the typological study of linguistic structures. In Hjelmslev's interpretation, there are no physical, psychological or other a priori principles that explain why languages are the way they are. Cross-linguistic similarities on the expression plane depend on a necessity to express meaning; conversely, cross-linguistic similarities on the content plane depend on the necessity to structure meaning potential according to the necessities of expression.

"The linguist must be equally interested in the similarity and in the difference between languages, two complementary sides of the same thing. The similarity between languages is their very structural principle; the difference between languages is the carrying out of that principle in *concreto*. Both the similarity and the difference between languages lie, then, in language and in languages themselves, in their internal structure; and no similarity or difference between languages rests on any factor outside language." – Louis Hjelmslev

# Compositional and combinatorial language

According to André Martinet's concept of *double articulation*, language is a double-levelled or doubly articulated system. In this context, 'articulation' means 'joining'. The first level of articulation involves minimally meaningful units (*monemes*: words or morphemes), while the second level consists of minimally distinct non-signifying units (phonemes). Owing to double articulation, it is possible to construct all necessary words of a language with a couple dozen phonic units. Meaning is associated with combinations of the non-meaningful units. The organisation of language into hierarchical inventories makes highly complex and therefore highly useful language possible:

- "We might imagine a system of communication in which a special cry would correspond to each given situations and these facts of experience, it will be clear that if such a system were to serve the same purpose as our languages, it would have to comprise so large a number of distinct signs that the memory of man would be incapable of storing it. A few thousand of such units as *tête, mal, ai, la*, freely combinable, enable us to communicate more things than could be done by millions of unarticulated cries." – André Martinet

Louis Hjelmslev's conception includes even more levels: phoneme, morpheme, lexeme, phrase, sentence and discourse. Building on the smallest meaningful and non-meaningful elements, *glossemes*,it is possible to generate an infinite number of productions:

- "When we compare the inventories yielded at the various stages of the deduction, their size will usually turn out to decrease as the procedure goes on. If the text is unrestricted, i.e., capable of being prolonged through constant addition of further parts ... it will be possible to register an unrestricted number of sentences." – Louis Hjelmslev

These notions are a continuation in a humanistictradition which considers language as a human invention. A similar idea is found in Port-Royal Grammar:

- "It remains for us to examine the spiritual element of speech ... this marvelous invention of composing from twenty-five or thirty sounds an infinite variety of words, which, although not having any resemblance in themselves to that which passes through our minds, nevertheless do not fail to reveal to others all of the secrets of the mind, and to make intelligible to others who cannot penetrate into the mind all that we conceive and all of the diverse movements of our souls." – Antoine Arnauld

## Interaction of meaning and form

Another way to approach structural explanation is from Saussure's concept of semiology (semiotics). Language is considered as arising from the interaction of form and meaning. Saussure's concept of the bilateral sign (signifier – signified) entails that the conceptual system is distinct from physical reality. For example, the spoken sign 'cat' is an association between the combination of the sounds [k], [æ] and

[t] and the concept of a cat, rather than with its referent (an actual cat). Language is thus considered a fully abstract system where each item in the conceptual inventory is associated with an expression; and these two levels define, organise and restrict each other.

Key concepts of the organisation of the phonemic versus the semantic system are those of opposition and distinctiveness. Each phoneme is distinct from other phonemes of the phonological system of a given language. The concepts of distinctiveness and markednesswere successfully used by the Prague Linguistic Circle to explain the phonemic organisation of languages, laying a ground for modern phonology as the study of the sound systems of languages, also borrowing from Wilhelm von Humboldt.

Likewise, each concept is distinct from all others in the conceptual system, and is defined in opposition with other concepts. Louis Hjelmslev laid the foundation of structural semantics with his idea that the content-level of language has a structure analogous to the level of expression. Structural explanation in the sense of how language shapes our understanding of the world has been widely used by the post-structuralists.

Structural linguist Lucien Tesnière, who invented dependency grammar, considered the relationship between meaning and form as conflicting due to a mathematical difference in how syntactic and semantic structure is organised. He used his concept of *antinomy* between syntax and semantics to elucidate the concept of a language as a solution to the communication problem. From his perspective, the two-dimensional semantic

dependency structure is necessarily forced into one-dimensional (linear) form. This causes the meaningful semantic arrangement to break into a largely arbitrary word ordering.

# Recent perceptions of structuralism

Those working in the generativist tradition often regard structuralist approaches as outdated and superseded. For example, Mitchell Marcus writes that structural linguistics was "fundamentally inadequate to process the full range of natural language". Holland writes that Chomsky had "decisively refuted Saussure". Similar views have been expressed by Jan Koster, Mark Turner, and other advocates of sociobiology.

Others however stress the continuing importance of Saussure's thought and structuralist approaches. Gilbert Lazard has dismissed the Chomskyan approach as passé while applauding a return to Saussurean structuralism as the only course by which linguistics can become more scientific. Matthews notes the existence of many "linguists who are structuralists by many of the definitions that have been proposed, but who would themselves vigorously deny that they are anything of the kind", suggesting a persistence of the structuralist paradigm.

# Effect of structuralist linguistics upon other disciplines

In the 1950s Saussure's ideas were appropriated by several prominent figures in Continental philosophy, anthropology, and from there were borrowed in literary theory, where they are used to interpret novels and other texts. However, several critics have charged that Saussure's ideas have been misunderstood or deliberately distorted by continental philosophers and literary theorists and are certainly not directly applicable to the textual level, which Saussure himself would have firmly placed within parole and so not amenable to his theoretical constructs.

# Theory of language

**Theory of language** is a topic from philosophy of language and theoretical linguistics. It has the goal of answering the questions “What is language?”; "Why do languages have the properties they have?"; or "What is the origin of language?".

Even though much of the research in linguistics is descriptive or prescriptive, there exists an underlying assumption that terminological and methodological choices reflect the researcher's opinion of language. Linguists are divided into different schools of thinking, with the nature–nurture debate as the main divide. Some linguistics conferences and journals are focussed on a specific theory of language, while others disseminate a variety of views.

Like in other human and social sciences, theories in linguistics can be divided into humanistic and sociobiological approaches. Same terms, for example 'rationalism', 'functionalism', 'formalism' and 'constructionism', are used with different meanings in different contexts.

# Humanistic theories

Humanistic theories consider people as having an agentive role in the social construction of language. Language is primarily seen as a sociocultural phenomenon. This tradition emphasises culture, nurture, creativity and diversity. A classical rationalist approach to language stems from the philosophy Age of Enlightenment. Francisco Sánchez de las Brozas and Antoine Arnauld believed that people had created language in a step-by-step process to serve their psychological need to communicate with each other. Thus, language is thought of as a rational human invention.

## Cultural–historical approaches

During the 19th century, when sociological questions remained under psychology, languages and language changewere thought of as arising from human psychology and the collective unconscious mind of the community, shaped by its history, as argued by Moritz Lazarus, HeymannSteinthal and Wilhelm Wundt. Advocates of Völkerpsychologie ('folk psychology') regarded language as Volksgeist; a social phenomenon conceived as the 'spirit of the nation'.

Wundt claimed that the human mind becomes organised according to the principles of syllogistic reasoning with social progress and education. He argued for a binary-branching model for the description of the mind, and syntax. Folk psychology was imported to North American linguistics by Franz Boas and Leonard Bloomfield who were the founders of a school of thought which was later nicknamed 'American structuralism'.

Folk psychology became associated with German nationalism, and after World War I Bloomfield apparently replaced Wundt's structural psychology with Albert Paul Weiss's behavioral psychology; although Wundtian notions remained elementary for his linguistic analysis. The Bloomfieldian school of linguistics was eventually reformed as a sociobiological approach by Noam Chomsky (see 'generative grammar' below).

Since generative grammar's popularity began to wane towards the end of the 20th century, there has been a new wave of cultural anthropological approaches to the language question sparking a modern debate on the relationship of language and culture. Participants include Daniel Everett, Jesse Prinz, Nicholas Evans and Stephen Levinson.

## Structuralism: a sociological–semiotic theory

The study of culture and language developed in a different direction in Europe where Émile Durkheim successfully separated sociology from psychology, thus establishing it as an autonomous science. Ferdinand de Saussure likewise argued

for the autonomy of linguistics from psychology. He created a semiotictheory which would eventually give rise to the movement in human sciences known as structuralism, followed by functionalism or functional structuralism, post-structuralism and other similar tendencies. The names structuralism and functionalism are derived from Durkheim's modification of Herbert Spencer's organicism which draws an analogy between social structures and the organs of an organism, each necessitated by its function.

Saussure approaches the essence of language from two sides. For the one, he borrows ideas from Steinthal and Durkheim, concluding that language is a 'social fact'. For the other, he creates a theory of language as a system in and for itself which arises from the association of concepts and words or expressions. Thus, language is a dual system of interactive sub-systems: a conceptual system and a system of linguistic forms. Neither of these can exist without the other because, in Saussure's notion, there are no (proper) expressions without meaning, but also no (organised) meaning without words or expressions. Language as a system does not arise from the physical world, but from the contrast between the concepts, and the contrast between the linguistic forms.

## Functionalism: language as a tool for communication

There was a shift of focus in sociology in the 1920s, from structural to functional explanation, or the adaptation of the social 'organism' to its environment. Post-Saussurean linguists, led by the Prague linguistic circle, began to study

the functional value of the linguistic structure, with communication taken as the primary function of language in the meaning 'task' or 'purpose'. These notions translated into an increase of interest in pragmatics, with a discourse perspective (the analysis of full texts) added to the multilayered interactive model of structural linguistics. This gave rise to functional linguistics.

## Formalism: language as a mathematical–semiotic system

Structural and formal linguistLouis Hjelmslev considered the systemic organisation of the bilateral linguistic system fully mathematical, rejecting the psychological and sociological aspect of linguistics altogether. He considered linguistics as the comparison of the structures of all languages using formal grammars – semantic and discourse structures included. Hjelmslev's idea is sometimes referred to as 'formalism'.

Although generally considered as a structuralist, Lucien Tesnière regarded meaning as giving rise to expression, but not vice versa, at least as regards the relationship between semantics and syntax. He considered the semantic plane as psychological, but syntax as being based on the necessity to break the two-dimensional semantic representation into linear form.

## Post-structuralism: language as a societal tool

The Saussurean idea of language as an interaction of the conceptual system and the expressive system was elaborated in philosophy, anthropology and other fields of human sciences by Claude Lévi-Strauss, Roland Barthes, Michel Foucault, Jacques Derrida, Julia Kristeva and many others. This movement was interested in the Durkheimian concept of language as a social fact or a rule-based code of conduct; but eventually rejected the structuralist idea that the individual cannot change the norm. Post-structuralists study how language affects our understanding of reality thus serving as a tool of shaping society.

## Language as an artificial construct

While the humanistic tradition stemming from 19th century Völkerpsychologie emphasises the unconscious nature of the social construction of language, some perspectives of post-structuralism and social constructionism regard human languages as man-made rather than natural. At this end of the spectrum, structural linguist Eugenio Coșeriu laid emphasis on the intentional construction of language. Daniel Everett has likewise approached the question of language construction from the point of intentionality and free will.

There were also some contacts between structural linguists and the creators of constructed languages. For example, Saussure's brother René de Saussure was an Esperanto

activist, and the French functionalist André Martinet served as director of the International Auxiliary Language Association.

# Sociobiological theories

In contrast to humanistic linguistics, sociobiological approaches consider language as a biological phenomena. Approaches to language as part of cultural evolution can be roughly divided into two main groups: genetic determinism which argues that languages stem from the human genome; and social Darwinism, as envisioned by August Schleicher and Max Müller, which applies principles and methods of evolutionary biology to linguistics. Because sociobiogical theories have been labelled as chauvinistic in the past, modern approaches, including Dual inheritance theory and memetics, aim to provide more sustainable solutions to the study of biology's role in language.

## Language as a genetically inherited phenomenon

# Strong version ('rationalism')

The role of genes in language formation has been discussed and studied extensively. Proposing generative grammar, Noam Chomsky argues that language is fully caused by a random genetic mutation, and that linguistics is the study of universal grammar, or the structure in question. Others, including Ray Jackendoff, point out that the innate language component

could be the result of a series of evolutionary adaptations; Steven Pinker argues that, because of these, people are born with a language instinct.

The random and the adaptational approach are sometimes referred to as formalism (or structuralism) and functionalism (or adaptationism), respectively, as a parallel to debates between advocates of structural and functional explanation in biology. Also known asbiolinguistics, the study of linguistic structures is parallelised with that of natural formations such as ferromagnetic droplets and botanic forms. This approach became highly controversial at the end of the 20th century due to a lack of empirical support for genetics as an explanation of linguistic structures.

More recent anthropological research aims to avoid genetic determinism. Behavioural ecology and dual inheritance theory, the study of gene–culture co-evolution, emphasise the role of culture as a human invention in shaping the genes, rather than vice versa. It is known, for example, that since early humans started developing their language, the process paved way for genetic changes that would affect the vocal tract.

## Weak version ('empiricism')

Some former generative grammarians argue that genes may nonetheless have an indirect effect on abstract features of language. This makes up yet another approach referred to as 'functionalism' which makes a weaker claim with respect to genetics. Instead of arguing for a specific innate structure, it is

suggested that human physiology and neurologicalorganisation may give rise to linguistic phenomena in a more abstract way.

Based on a comparison of structures from multiple languages, John A. Hawkins suggests that the brain, as a syntactic parser, may find it easier to process some word orders than others, thus explaining their prevalence. This theory remains to be confirmed by psycholinguistic studies.

Conceptual metaphor theory from George Lakoff's cognitive linguistics hypothesises that people have inherited from lower animals the ability for deductive reasoning based on visual thinking, which explains why languages make so much use of visual metaphors.

## Languages as species

It was thought in early evolutionary biology that languages and speciescan be studied according to the same principles and methods. The idea of languages and cultures as fighting for living space became highly controversial as it was accused of being a pseudoscience that caused two world wars, and social Darwinism was banished from humanities by 1945. In the concepts of Schleicher and Müller, both endorsed by Charles Darwin, languages could be either organisms or populations.

A neo-Darwinian version of this idea was introduced as memetics by Richard Dawkins in 1976. In this thinking, ideas and cultural units, including words, are compared to viruses or replicators. Although meant as a softer alternative to genetic determinism, memetics has been widely discredited as pseudoscience, and it has failed to establish itself as a

recognised field of scientific research. The language–species analogy nonetheless continues to enjoy popularity in linguistics and other human sciences. Since the 1990s there have been numerous attempts to revive it in various guises. As JaminPelkey explains,

"Theorists who explore such analogies usually feel obliged to pin language to some specific sub-domain of biotic growth. William James selects “zoölogical evolution”, William Croft prefers botanical evolution, but most theorists zoom in to more microbiotic levels – some claiming that linguistic phenomena are analogous to the cellular level and others arguing for the genetic level of biotic growth. For others, language is a parasite; for others still, language is a virus ... The disagreements over grounding analogies do not stop here."

Like many other approaches to linguistics, these, too, are collectively called 'functionalism'. They include various frameworks of usage-based linguistics, language as a complex adaptive system, construction grammar, emergent linguistics, and others.

Chapter 7

# Stubs

# Linguistic conservatism

In linguistics, a **conservative** form, variety, or modality is one that has changed relatively little over its history, or which is relatively resistant to change. It is the opposite of **innovative** or **advanced** forms or varieties, which have undergone relatively larger or more recent changes. On the other hand, an *archaic* language is chronologically old.

A conservative linguistic form, such as a word, is one that remains closer to an older form from which it evolved, relative to cognate forms from the same source. For example, the Spanish word *caro* and the French word *cher* both evolved from the Latin word *cārum*. The Spanish word, which is more similar to the common ancestor, is more conservative than its French cognate.

A language or language varietyis said to be conservative if it has fewer innovations (in other words, more conservative forms) than related varieties do. For example, Icelandic is, in some aspects, more similar to Old Norse than other languages that evolved from Old Norse, including Danish, Norwegian, or Swedish, while Sardinian (especially the Nuorese dialects) is regarded by many linguists to be the most conservative Romance language. In fact, recent studies regarding the stability of modern Icelandic appear to confirm its status as

"stable". Therefore, Icelandic and Sardinian are considered relatively conservative languages. Likewise, some dialects of a language may be more conservative than others. Standard varieties, for example, tend to be more conservative than nonstandard varieties, since education and codification in writing tend to retard change.

Writing is generally said to be more conservative than speech since written forms generally change more slowly than spoken language does. That helps explain inconsistencies in writing systems such as that of English; since the spoken language has changed relatively more than has the written language, the match between spelling and pronunciation is inconsistent. (See Great Vowel Shift.)

A language may be conservative in one respect while simultaneously innovative in another. Bulgarian and Macedonian, closely related Slavic languages, are innovative in the grammar of their nouns, having dropped nearly all vestiges of the complex Slavic case system; at the same time, they are highly conservative in their verbal system, which has been greatly simplified in most other Slavic languages. English, which is one of the more innovative Germanic languages in most respects (vocabulary, inflection, vowel phonology, syntax), is nevertheless conservative in its consonant phonology, retaining sounds such as (most notably) /θ/ and /ð/ (*th*), which remain only in English, Icelandic and Scots.

Conservative languages are often thought of as being more complex grammatically (or at least morphologically) than innovative languages. That is largely true for Indo-European languages, where the parent language had an extremely

complex morphology and the dominant pattern of language change has been simplified. On the other hand, a number of Arabic varieties commonly considered innovative, such as Egyptian Arabic, have developed a complex agglutinative system of verbal morphology out of the simpler system of Classical Arabic.

In the 6th century AD, Classical Arabic was a conservative Semitic language compared with Classical Syriac, which was spoken at the same time; Classical Arabic strongly resembles reconstructed Proto-Semitic, and Syriac has changed much more. Compared to closely related modern Northeastern Neo-Aramaic, which is not necessarily directly descended from it, Classical Syriac is still a highly archaic language form.

Georgian has changed remarkably little since the Old Georgian period (the 4th/5th century AD).

A roughly analogous concept in biology is living fossil.

# Principle of compositionality

In semantics, mathematical logic and related disciplines, the **principle of compositionality** is the principle that the meaning of a complex expression is determined by the meanings of its constituent expressions and the rules used to combine them. This principle is also called **Frege's principle**, because GottlobFrege is widely credited for the first modern formulation of it. The principle was never explicitly stated by Frege, and it was arguably already assumed by George Boole decades before Frege's work.

The principle of compositionality is highly debated in linguistics, and among its most challenging problems there are the issues of contextuality, the non-compositionality of idiomatic expressions, and the non-compositionality of quotations.

# History

Discussion of compositionality started to appear at the beginning of the 19th century, during which it was debated whether what was most fundamental in language was compositionality or contextuality, and compositionality was usually preferred. Frege (1848-1925) never adhered to the principle of compositionality as it is known today, and the first to explicitly formulate it was Freges' student Rudolf Carnap in 1947.

# Overview

The principle of compositionality states that in a meaningful sentence, if the lexical parts are taken out of the sentence, what remains will be the rules of composition. Take, for example, the sentence "Socrates was a man". Once the meaningful lexical items are taken away—"Socrates" and "man"—what is left is the pseudo-sentence, "S was a M". The task becomes a matter of describing what the connection is between S and M.

Among the most prominent linguistic problems that challenge the principle of compositionality are the issues of

contextuality, the non compositionality of idiomatic expressions, and the non compositionality of quotations.

It is frequently taken to mean that every operation of the syntax should be associated with an operation of the semantics that acts on the meanings of the constituents combined by the syntactic operation. As a guideline for constructing semantic theories, this is generally taken, as in the influential work on the philosophy of language by Donald Davidson, to mean that every construct of the syntax should be associated by a clause of the T-schema with an operator in the semantics that specifies how the meaning of the whole expression is built from constituents combined by the syntactic rule. In some general mathematical theories (especially those in the tradition of Montague grammar), this guideline is taken to mean that the interpretation of a language is essentially given by a homomorphism between an algebra of syntactic representations and an algebra of semantic objects.

The principle of compositionality also exists in a similar form in the compositionality of programming languages.

# Critiques

The principle of compositionality has been the subject of intense debate. Indeed, there is no general agreement as to how the principle is to be interpreted, although there have been several attempts to provide formal definitions of it.

Scholars are also divided as to whether the principle should be regarded as a factual claim, open to empirical testing; an analytic truth, obvious from the nature of language and

meaning; or a methodological principle to guide the development of theories of syntax and semantics. The Principle of Compositionality has been attacked in all three spheres, although so far none of the criticisms brought against it have been generally regarded as compelling. Most proponents of the principle, however, make certain exceptions for idiomatic expressions in natural language.

The principle of compositionality usually holds when only syntactic factors play in the increased complexity of sentence processing, while it becomes more problematic and questionable when the complexity increase is due to sentence or discourse context, semantic memory, or sensory cues. Among the problematic phenomena for traditional theories of compositionality is that of logical metonymy, which has been studied at least since the mid 1990s by linguists James Pustejovsky and Ray Jackendoff. Logical metonymies are sentences like *John began the book*, where the verb *to begin* requires (subcategorizes) an event as its argument, but in a logical metonymy an object (i.e. *the book*) is found instead, and this forces to interpret the sentence by inferring an implicit event ("reading", "writing", or other prototypical actions performed on a book). The problem for compositionality is that the meaning of reading or writing is not present in the words of the sentence, neither in "begin" nor in "book".

Further, in the context of the philosophy of language, the principle of compositionality does not explain all of meaning. For example, you cannot infer sarcasm purely on the basis of words and their composition, yet a phrase used sarcastically means something completely different from the same phrase uttered straightforwardly. Thus, some theorists argue that the

principle has to be revised to take into account linguistic and extra linguistic context, which includes the tone of voice used, common ground between the speakers, the intentions of the speaker, and so on.

# Autonomy of syntax

In linguistics, the **autonomy of syntax** is the assumption that syntax is arbitrary and self-contained with respect to meaning, semantics, pragmatics, discourse function, and other factors external to language. The autonomy of syntax is advocated by linguistic formalists, and in particular by generative linguistics, whose approaches have hence been called autonomist linguistics.

The autonomy of syntax is at the center of the debates between formalist and functionalist linguistics, and since the 1980s research has been conducted on the syntax–semantics interface within functionalist approaches, aimed at finding instances of semantically determined syntactic structures, to disprove the formalist argument of the autonomy of syntax.

The principle of iconicityis contrasted, for some scenarios, with that of the autonomy of syntax. The weaker version of the argument for the autonomy of syntax (or that for the autonomy of grammar), includes only for the principle of arbitrariness, while the stronger version includes the claim of self-containedness. The principle of arbitrariedness of syntax is actually accepted by most functionalist linguist, and the real dispute between functionalist and generativists is on the claim of self-containedness of grammar or syntax.

# History

The assumption of the autonomy of syntax can be traced back to the neglect of the study of semantics by American structuralists like Leonard Bloomfield and Zellig Harris in the 1940s, which was based on a neo-positivistanti-psychologist stance, according to which since it's presumably impossibile to study how the brain works, linguists should ignore all cognitive and psychological aspects of language and focus on the only objective data, that is how language appears in its exterior form. This paralleled the distinction between the two approaches in psychology, behaviorism, which was the dominant approach up until the 1940s, and cognitivism.

Over the decades, multiple instances have been found of cases in which syntactic structures are actually determined or influenced by semantic traits, and some formalists and generativists have reacted to that by shrinking those parts of semantics that they consider autonomous. Over the decades, in the changes that Noam Chomsky has made to his generative formulation, there has been a shift from a claim for the autonomy of syntax to one for the autonomy of grammar.

# Etymology

**Etymology** (/ˌɛtɪˈmɒlədʒi/) is the study of the history of words. By extension, the etymology of a word means its origin and development throughout history.

For languages with a long written history, etymologists make use of texts, and texts about the language, to gather knowledge about how words were used during earlier periods, how they developed in meaning and form, or when and how they entered the language. Etymologists also apply the methods of comparative linguistics to reconstruct information about forms that are too old for any direct information to be available.

By analyzing related languages with a technique known as the comparative method, linguists can make inferences about their shared parent language and its vocabulary. In this way, word roots in European languages, for example, can be traced all the way back to the origin of the Indo-Europeanlanguage family.

Even though etymological research originally grew from the philological tradition, much current etymological research is done on language families where little or no early documentation is available, such as Uralic and Austronesian.

# Etymology

The word *etymology* derives from the Greek word ἐτυμολογία (*etumología*), itself from ἔτυμον (*étumon*), meaning "true sense or sense of a truth", and the suffix *-logia*, denoting "the study of".

The term *etymon* refers to a word or morpheme (e.g., stem or root) from which a later word or morpheme derives. For example, the Latin word *candidus*, which means "white", is the etymon of English *candid*. Relationships are often less transparent, however. English place names such as

Winchester, Gloucester, Tadcaster share in different modern forms a suffixed etymon that was once meaningful, Latin *castrum* 'fort'.

# Methods

Etymologists apply a number of methods to study the origins of words, some of which are:

- Philological research. Changes in the form and meaning of the word can be traced with the aid of older texts, if such are available.
- Making use of dialectological data. The form or meaning of the word might show variations between dialects, which may yield clues about its earlier history.
- The comparative method. By a systematic comparison of related languages, etymologists may often be able to detect which words derive from their common ancestor language and which were instead later borrowed from another language.
- The study of semantic change. Etymologists must often make hypotheses about changes in the meaning of particular words. Such hypotheses are tested against the general knowledge of semantic shifts. For example, the assumption of a particular change of meaning may be substantiated by showing that the same type of change has occurred in other languages as well.

# Types of word origins

Etymological theory recognizes that words originate through a limited number of basic mechanisms, the most important of which are language change, borrowing (i.e., the adoption of "loanwords" from other languages); word formation such as derivation and compounding; and onomatopoeia and sound symbolism (i.e., the creation of imitative words such as "click" or "grunt").

While the origin of newly emerged words is often more or less transparent, it tends to become obscured through time due to sound change or semantic change. Due to sound change, it is not readily obvious that the English word *set*is related to the word *sit* (the former is originally a causative formation of the latter). It is even less obvious that *bless* is related to *blood* (the former was originally a derivative with the meaning "to mark with blood").

Semantic change may also occur. For example, the English word *bead* originally meant "prayer". It acquired its modern meaning through the practice of counting the recitation of prayers by using beads.

# English language

English derives from Old English (sometimes referred to as Anglo-Saxon), a West Germanic variety, although its current vocabulary includes words from many languages. The Old English roots may be seen in the similarity of numbers in

English and German, particularly *seven/sieben,eight/acht,nine/neun,* and *ten/zehn.*Pronouns are also cognate: *I/mine/me and ich/mein/mich*; *thou/thine/thee and du/dein/dich*; *we/wir and us/uns*; *she/sie; your/ihr.* However, language change has eroded many grammatical elements, such as the noun case system, which is greatly simplified in modern English. Certain elements of vocabulary are borrowed from French and other Romance languages, but most of the common words used in English are of Germanic origin.

When the Normans conquered England in 1066 (see Norman Conquest), they brought their Norman language with them. During the Anglo-Norman period, which united insular and continental territories, the ruling class spoke Anglo-Norman, while the peasants spoke the vernacular English of the time. Anglo-Norman was the conduit for the introduction of French into England, aided by the circulation of Langue d'oïl literature from France.

This led to many paired words of French and English origin. For example, *beef*is related, through borrowing, to modern French *bœuf,veal* to *veau,pork* to *porc,* and *poultry* to *poulet.* All these words, French and English, refer to the meat rather than to the animal. Words that refer to farm animals, on the other hand, tend to be cognates of words in other Germanic languages. For example, *swine/Schwein,cow/Kuh,calf/Kalb,* and *sheep/Schaf.* The variant usage has been explained by the proposition that it was the Norman rulers who mostly ate meat (an expensive commodity) and the Anglo-Saxons who farmed the animals. This explanation has passed into common folklore but has been disputed.

# Assimilation of foreign words

English has proved accommodating to words from many languages. Scientific terminology, for example, relies heavily on words of Latin and Greek origin, but there are a great many non-scientific examples. Spanish has contributed many words, particularly in the southwestern United States. Examples include *buckaroo*, *alligator*, *rodeo*, *savvy*, and states' names such as *Colorado* and *Florida*. *Albino*, *palaver*, *lingo*, *verandah*, and *coconut* from Portuguese; *diva* and *prima donna* from Italian. Modern French has contributed *café*, *cinema*, *naive*, *nicotine* and many more.

*Smorgasbord*, *slalom*, and *ombudsman* are from Swedish, Norwegian and Danish; *sauna* from Finnish; *adobe*, *alcohol*, *algebra*, *algorithm*, *apricot*, *assassin*, *caliber*, *cotton*, *hazard*, *jacket*, *jar*, *julep*, *mosque*, *Muslim*, *orange*, *safari*, *sofa*, and *zero* from Arabic (often via other languages); *behemoth*, *hallelujah*, *Satan*, *jubilee*, and *rabbi* from Hebrew; *taiga*, *steppe*, *Bolshevik*, and *sputnik* from Russian.

*Bandanna*, *bungalow*, *dungarees*, *guru*, *karma*, and *pundit* come from Urdu, Hindi and ultimately Sanskrit; *curry* from Tamil; *honcho*, *sushi*, and *tsunami* from Japanese; *dim sum*, *gung ho*, *kowtow*, *kumquat* and *typhoon* from Cantonese. *Kampong* and *amok* are from Malay; and *boondocks* from the Tagalog word for hills or mountains, *bundok*. *Ketchup* derives from one or more South-East Asia and East Indies words for fish sauce or soy sauce, likely by way of Chinese, though the precise path is unclear: Malay*kicap*, Indonesian*kecap*, Chinese Min Nan*kê-chiap* and cognates in other Chinese dialects.

Surprisingly few loanwords, however, come from other languages native to the British Isles. Those that exist include *coracle*, *cromlech* and (probably) *flannel*, *gull* and *penguin* from Welsh; *galore* and *whisky* from Scottish Gaelic; *phoney*, *trousers*, and *Tory* from Irish; and *eerie* and *canny* from Scots (or related Northern English dialects).

Many Canadian English and American English words (especially but not exclusively plant and animal names) are loanwords fromIndigenous American languages, such as *barbecue*, *bayou*, *chili*, *chipmunk*, *hooch*, *hurricane*, *husky*, *mesquite*, *opossum*, *pecan*, *squash*, *toboggan*, and *tomato*.

# History

The search for meaningful origins for familiar or strange words is far older than the modern understanding of linguistic evolution and the relationships of languages, which began no earlier than the 18th century. From Antiquity through the 17th century, from *Pāṇini* to Pindar to Sir Thomas Browne, etymology had been a form of witty wordplay, in which the supposed origins of words were creatively imagined to satisfy contemporary requirements; for example, the Greek poetPindar (born in approximately 522 BCE) employed inventive etymologies to flatter his patrons. Plutarch employed etymologies insecurely based on fancied resemblances in sounds. Isidore of Seville's *Etymologiae* was an encyclopedic tracing of "first things" that remained uncritically in use in Europe until the sixteenth century. *Etymologicumgenuinum* is a grammaticalencyclopedia edited at Constantinople in the ninth century, one of several similar Byzantine works. The

thirteenth-century *LegendaAurea*, as written by Jacobus de Vorgagine, begins each *vita* of a saint with a fanciful excursus in the form of an etymology.

## Ancient Sanskrit

The Sanskrit linguists and grammarians of ancient India were the first to make a comprehensive analysis of linguistics and etymology. The study of Sanskrit etymology has provided Western scholars with the basis of historical linguistics and modern etymology. Four of the most famous Sanskrit linguists are:

- Yaska (c. 6th–5th centuries BCE)
- *Pāĝini* (c. 520–460 BCE)
- *Kātyāyana* (2nd century BCE)
- *Patañjali* (2nd century BCE)

These linguists were not the earliest Sanskrit grammarians, however. They followed a line of ancient grammarians of Sanskrit who lived several centuries earlier like Sakatayana of whom very little is known. The earliest of attested etymologies can be found in Vedic literature in the philosophical explanations of the *Brahmanas*, *Aranyakas*, and *Upanishads*.

The analyses of Sanskrit grammar done by the previously mentioned linguists involved extensive studies on the etymology (called *Nirukta* or *Vyutpatti* in Sanskrit) of Sanskrit words, because the ancient Indians considered sound and speech itself to be sacred and, for them, the words of the sacred *Vedas* contained deep encoding of the mysteries of the soul and God.

## Ancient Greco-Roman

One of the earliest philosophical texts of the Classical Greek period to address etymology was the Socratic dialogue *Cratylus* (c. 360 BCE) by Plato. During much of the dialogue, Socrates makes guesses as to the origins of many words, including the names of the gods. In his OdesPindar spins complimentary etymologies to flatter his patrons. Plutarch (*Life of NumaPompilius*) spins an etymology for *pontifex*, while explicitly dismissing the obvious, and actual "bridge-builder":

The priests, called Pontifices.... have the name of Pontifices from *potens*, powerful because they attend the service of the gods, who have power and command overall. Others make the word refer to exceptions of impossible cases; the priests were to perform all the duties possible; if anything lays beyond their power, the exception was not to be cavilled. The most common opinion is the most absurd, which derives this word from pons, and assigns the priests the title of bridge-makers. The sacrifices performed on the bridge were amongst the most sacred and ancient, and the keeping and repairing of the bridge attached, like any other public sacred office, to the priesthood.

## Medieval

Isidore of Seville compiled a volume of etymologies to illuminate the triumph of religion. Each saint's legend in Jacob de Voragine's *LegendaAurea* begins with an etymological discourse on the saint's name:

Lucy is said of light, and light is beauty in beholding, after that S. Ambrose saith: The nature of light is such, she is gracious in beholding, she spreadeth over all without lying down, she passeth in going right without crooking by right long line; and it is without dilation of tarrying, and therefore it is showed the blessed Lucy hath beauty of virginity without any corruption; essence of charity without disordinate love; rightful going and devotion to God, without squaring out of the way; right long line by continual work without negligence of slothful tarrying. In Lucy is said, the way of light.

## Modern era

Etymology in the modern sense emerged in the late 18th-century European academia, within the context of the wider "Age of Enlightenment," although preceded by 17th century pioneers such as Marcus Zuerius van Boxhorn, Gerardus Vossius, Stephen Skinner, Elisha Coles, and William Wotton. The first known systematic attempt to prove the relationship between two languages on the basis of similarity of grammar and lexicon was made in 1770 by the Hungarian, JánosSajnovics, when he attempted to demonstrate the relationship between Sami and Hungarian (work that was later extended to the whole Finno-Ugric language family in 1799 by his fellow countryman, Samuel Gyarmathi).

The origin of modern historical linguisticsis often traced to Sir William Jones, a Welsh philologist living in India, who in 1782 observed the genetic relationship between Sanskrit, Greek and Latin. Jones published his *The Sanscrit Language* in 1786, laying the foundation for the field of Indo-European linguistics.

The study of etymology in Germanic philology was introduced by Rasmus Christian Rask in the early 19th century and elevated to a high standard with the *German Dictionary* of the Brothers Grimm. The successes of the comparative approach culminated in the Neogrammarian school of the late 19th century. Still in the 19th century, German philosopher Friedrich Nietzsche used etymological strategies (principally and most famously in *On the Genealogy of Morals*, but also elsewhere) to argue that moral values have definite historical (specifically, cultural) origins where modulations in meaning regarding certain concepts (such as "good" and "evil") show how these ideas had changed over time—according to which value-system appropriated them. This strategy gained popularity in the 20th century, and philosophers, such as Jacques Derrida, have used etymologies to indicate former meanings of words to de-center the "violent hierarchies" of Western philosophy.

# Notable etymologists

- Ernest Klein (1899-1983), Hungarian-born Romanian-Canadian linguist, etymologist
- Marko Snoj (born 1959), Indo-Europeanist, Slavist, Albanologist, lexicographer, and etymologist
- Anatoly Liberman(born 1937), linguist, medievalist, etymologist, poet, translator of poetry and literary critic
- Michael Quinion (born c. 1943)