

# Explaining Theoretical Language and Linguistics

Micheal Graves





**EXPLAINING  
THEORETICAL LANGUAGE  
AND LINGUISTICS**



# **EXPLAINING THEORETICAL LANGUAGE AND LINGUISTICS**

Micheal Graves



Explaining Theoretical Language and Linguistics  
by Micheal Graves

Copyright© 2022 BIBLIOTEX

[www.bibliotex.com](http://www.bibliotex.com)

All rights reserved. No part of this book may be reproduced or used in any manner without the prior written permission of the copyright owner, except for the use brief quotations in a book review.

To request permissions, contact the publisher at [info@bibliotex.com](mailto:info@bibliotex.com)

Ebook ISBN: 9781984665348



Published by:

Bibliotex

Canada

Website: [www.bibliotex.com](http://www.bibliotex.com)

# Contents

Chapter 1	Linguistics: An Overview .....	1
Chapter 2	Applied Linguistics .....	32
Chapter 3	Language Revitalization .....	105
Chapter 4	Internet Linguistics .....	138
Chapter 5	Forensic Linguistics .....	172





## Chapter 1

# Linguistics: An Overview

**Linguistics** is the scientific study of language. It encompasses the analysis of every aspect of language, as well as the methods for studying and modelling them.

The traditional areas of linguistic analysis include phonetics, phonology, morphology, syntax, semantics, and pragmatics. Each of these areas roughly corresponds to phenomena found in human linguistic systems: sounds (and gesture, in the case of signed languages), minimal units (words, morphemes), phrases and sentences, and meaning and use.

Linguistics studies these phenomena in diverse ways and from various perspectives. Theoretical linguistics (including traditional descriptive linguistics) is concerned with building models of these systems, their parts (ontologies), and their combinatorics. Psycholinguistics builds theories of the processing and production of all these phenomena. These phenomena may be studied synchronically or diachronically (through history), in monolinguals or polyglots, in children or adults, as they are acquired or statically, as abstract objects or as embodied cognitive structures, using texts (corpora) or through experimental elicitation, by gathering data mechanically, through fieldwork, or through introspective judgment tasks. Computational linguistics implements theoretical constructs to parse or produce natural language or homologues. Neurolinguistics investigates linguistic phenomena by experiments on actual brain responses involving linguistic stimuli.

Linguistics is related to philosophy of language, stylistics and rhetoric, semiotics, lexicography, and translation.

## **Major subdisciplines**

### **Historical linguistics**

Historical linguistics is the study of language change, particularly with regards to a specific language or a group of languages. Western trends in historical linguistics date back to roughly the late 18th century, when the discipline grew out of philology (the study of ancient texts and antique documents).

Historical linguistics emerged as one of the first few sub-disciplines in the field, and was most widely practiced during the late 19th century. Despite a shift in focus in the twentieth century towards formalism and generative grammar, which studies the universal properties of language, historical research today still remains a significant field of linguistic inquiry. Subfields of the discipline include language change and grammaticalisation.

Historical linguistics studies language change either diachronically (through a comparison of different time periods in the past and present) or in a synchronic manner (by observing developments between different variations that exist within the current linguistic stage of a language).

At first, historical linguistics served as the cornerstone of comparative linguistics, which involves a study of the relationship between different languages. During this time,

scholars of historical linguistics were only concerned with creating different categories of language families, and reconstructing prehistoric proto languages by using the comparative method and the method of internal reconstruction. Internal reconstruction is the method by which an element that contains a certain meaning is re-used in different contexts or environments where there is a variation in either sound or analogy.

The reason for this had been to describe well-known Indo-European languages, many of which used to have long written histories. Scholars of historical linguistics also studied Uralic languages, another European language family for which very little written material existed back then. After this, there was significant work that followed on the corpora of other languages too, such as that of the Austronesian languages as well as of Native American language families.

The above approach of comparativism in linguistics is now, however, only a small part of the much broader discipline called historical linguistics. The comparative study of specific Indo-European languages is considered a highly specialised field today, while comparative research is carried out over the subsequent internal developments in a language. In particular, it is carried out over the development of modern standard varieties of languages, or over the development of a language from its standardised form to its varieties.

For instance, some scholars also undertook a study attempting to establish super-families, linking, for example, Indo-European, Uralic, and other language families to Nostratic. While these attempts are still not widely accepted as credible

methods, they provide necessary information to establish relatedness in language change, something that is not easily available as the depth of time increases. The time-depth of linguistic methods is generally limited, due to the occurrence of chance word resemblances and variations between language groups, but a limit of around 10,000 years is often assumed for the functional purpose of conducting research. Difficulty also exists in the dating of various proto languages. Even though several methods are available, only approximate results can be obtained in terms of arriving at dates for these languages.

Today, with a subsequent re-development of grammatical studies, historical linguistics studies the change in language on a relational basis between dialect to dialect during one period, as well as between those in the past and the present period, and looks at evolution and shifts taking place morphologically, syntactically, as well as phonetically.

## **Syntax and morphology**

Syntax and morphology are branches of linguistics concerned with the order and structure of meaningful linguistic units such as words and morphemes. Syntacticians study the rules and constraints that govern how speakers of a language can organize words into sentences. Morphologists study similar rules for the order of morphemes—sub-word units such as prefixes and suffixes—and how they may be combined to form words.

While words, along with clitics, are generally accepted as being the smallest units of syntax, in most languages, if not all,

many words can be related to other words by rules that collectively describe the grammar for that language. For example, English speakers recognize that the words *dog* and *dogs* are closely related, differentiated only by the plurality morpheme "-s", only found bound to noun phrases. Speakers of English, a fusional language, recognize these relations from their innate knowledge of English's rules of word formation. They infer intuitively that *dog* is to *dogs* as *cat* is to *cats*; and, in similar fashion, *dog* is to *dog catcher* as *dish* is to *dishwasher*. By contrast, Classical Chinese has very little morphology, using almost exclusively unbound morphemes ("free" morphemes) and depending on word order to convey meaning. (Most words in modern Standard Chinese ["Mandarin"], however, are compounds and most roots are bound.) These are understood as grammars that represent the morphology of the language. The rules understood by a speaker reflect specific patterns or regularities in the way words are formed from smaller units in the language they are using, and how those smaller units interact in speech. In this way, morphology is the branch of linguistics that studies patterns of word formation within and across languages and attempts to formulate rules that model the knowledge of the speakers of those languages.

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

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

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

## **Semantics and pragmatics**

Semantics and pragmatics are branches of linguistics concerned with meaning. These subfields have traditionally been divided according to aspects of meaning thought to arise from the grammar versus linguistic and social context. Semantics in this conception is concerned with grammatical and lexical meanings and pragmatics concerned with meaning in context. The framework of formal semantics studies the denotations of sentences and the way they are composed from the meanings of their constituent expressions. Formal semantics draws heavily on philosophy of language and uses formal tools from logic and computer science. Cognitive semantics ties linguistic meaning to general aspects of cognition, drawing on ideas from cognitive science such as prototype theory.

Pragmatics encompasses phenomena such as speech acts, implicature, and talk in interaction. Unlike semantics, which examines meaning that is conventional or "coded" in a given

language, pragmatics studies how the transmission of meaning depends not only on structural and linguistic knowledge (grammar, lexicon, etc.) of the speaker and listener but also on the context of the utterance, any pre-existing knowledge about those involved, the inferred intent of the speaker, and other factors. In that respect, pragmatics explains how language users are able to overcome apparent ambiguity since meaning relies on the manner, place, time, etc. of an utterance.

## **Phonetics and phonology**

Phonetics and phonology are branches of linguistics concerned with sounds (or the equivalent aspects of sign languages). Phonetics is largely concerned with the physical aspects of sounds such as their articulation, acoustics, production, and perception. Phonology is concerned with the linguistic abstractions and categorizations of sounds.

## **Typology**

- Linguistic typology (or language typology) is a field of linguistics that studies and classifies languages according to their structural features. Its aim is to describe and explain the common properties and the structural diversity of the world's languages. Its subdisciplines include, but are not limited to: qualitative typology, which deals with the issue of comparing languages and within-language variance; quantitative typology, which deals with the distribution of structural patterns in the world's languages; theoretical typology, which explains these

distributions; syntactic typology, which deals with word order, word form, word grammar and word choice; and lexical typology, which deals with language vocabulary.

## **Language varieties**

Languages exist on a wide continuum of conventionalization with blurry divisions between concepts such as dialects and languages. Languages can undergo internal changes which lead to the development of subvarieties such as linguistic registers, accents, and dialects. Similarly, languages can undergo changes caused by contact with speakers of other languages, and new language varieties may be born from these contact situations through the process of language genesis.

### **Contact varieties**

Contact varieties such as pidgins and creoles are language varieties that often arise in situations of sustained contact between communities that speak different languages. Pidgins are language varieties with limited conventionalization where ideas are conveyed through simplified grammars that may grow more complex as linguistic contact continues. Creole languages are language varieties similar to pidgins but with greater conventionalization and stability. As children grow up in contact situations, they may learn a local pidgin as their native language. Through this process of acquisition and transmission, new grammatical features and lexical items are created and introduced to fill gaps in the pidgin eventually developing into a complete language.



Not all language contact situations result in the development of a pidgin or creole, and researchers have studied the features of contact situations that make contact varieties more likely to develop. Often these varieties arise in situations of colonization and enslavement, where power imbalances prevent the contact groups from learning the other's language but sustained contact is nevertheless maintained. The subjugated language in the power relationship is the substrate language, while the dominant language serves as the superstrate. Often the words and lexicon of a contact variety come from the superstrate, making it the lexifier, while grammatical structures come from the substrate, but this is not always the case.

## **Dialect**

A dialect is a variety of language that is characteristic of a particular group among the language's speakers. The group of people who are the speakers of a dialect are usually bound to each other by social identity. This is what differentiates a dialect from a register or a discourse, where in the latter case, cultural identity does not always play a role. Dialects are speech varieties that have their own grammatical and phonological rules, linguistic features, and stylistic aspects, but have not been given an official status as a language. Dialects often move on to gain the status of a language due to political and social reasons. Other times, dialects remain marginalized, particularly when they are associated with marginalized social groups. Differentiation amongst dialects (and subsequently, languages) is based upon the use of grammatical rules, syntactic rules, and stylistic features, though not always on lexical use or vocabulary. The popular

saying that "a language is a dialect with an army and navy" is attributed as a definition formulated by Max Weinreich.

"We may as individuals be rather fond of our own dialect. This should not make us think, though, that it is actually any better than any other dialect. Dialects are not good or bad, nice or nasty, right or wrong – they are just different from one another, and it is the mark of a civilised society that it tolerates different dialects just as it tolerates different races, religions and sexes."

## **Standard language**

When a dialect is documented sufficiently through the linguistic description of its grammar, which has emerged through the consensual laws from within its community, it gains political and national recognition through a country or region's policies. That is the stage when a language is considered a standard variety, one whose grammatical laws have now stabilised from within the consent of speech community participants, after sufficient evolution, improvisation, correction, and growth. The English language, besides perhaps the French language, may be examples of languages that have arrived at a stage where they are said to have become standard varieties.

## **Relativity**

As constructed popularly through the Sapir-Whorf hypothesis, relativists believe that the structure of a particular language is capable of influencing the cognitive patterns through which a

person shapes his or her world view. Universalists believe that there are commonalities between human perception as there is in the human capacity for language, while relativists believe that this varies from language to language and person to person. While the Sapir-Whorf hypothesis is an elaboration of this idea expressed through the writings of American linguists Edward Sapir and Benjamin Lee Whorf, it was Sapir's student Harry Hoijer who termed it thus. The 20th century German linguist Leo Weisgerber also wrote extensively about the theory of relativity. Relativists argue for the case of differentiation at the level of cognition and in semantic domains. The emergence of cognitive linguistics in the 1980s also revived an interest in linguistic relativity. Thinkers like George Lakoff have argued that language reflects different cultural metaphors, while the French philosopher of language Jacques Derrida's writings, especially about deconstruction, have been seen to be closely associated with the relativist movement in linguistics, for which he was heavily criticized in the media at the time of his death.

## **Structures**

Linguistic structures are pairings of meaning and form. Any particular pairing of meaning and form is a Saussurean sign. For instance, the meaning "cat" is represented worldwide with a wide variety of different sound patterns (in oral languages), movements of the hands and face (in sign languages), and written symbols (in written languages). Linguistic patterns have proven their importance for the knowledge engineering field especially with the ever-increasing amount of available data.

Linguists focusing on structure attempt to understand the rules regarding language use that native speakers know (not always consciously). All linguistic structures can be broken down into component parts that are combined according to (sub)conscious rules, over multiple levels of analysis. For instance, consider the structure of the word "tenth" on two different levels of analysis. On the level of internal word structure (known as morphology), the word "tenth" is made up of one linguistic form indicating a number and another form indicating ordinality. The rule governing the combination of these forms ensures that the ordinality marker "th" follows the number "ten." On the level of sound structure (known as phonology), structural analysis shows that the "n" sound in "tenth" is made differently from the "n" sound in "ten" spoken alone. Although most speakers of English are consciously aware of the rules governing internal structure of the word pieces of "tenth", they are less often aware of the rule governing its sound structure. Linguists focused on structure find and analyze rules such as these, which govern how native speakers use language.

## **Grammar**

Grammar is a system of rules which governs the production and use of utterances in a given language. These rules apply to sound as well as meaning, and include componential subsets of rules, such as those pertaining to phonology (the organisation of phonetic sound systems), morphology (the formation and composition of words), and syntax (the formation and composition of phrases and sentences). Modern frameworks that deal with the principles of grammar include

structural and functional linguistics, and generative linguistics.

Sub-fields that focus on a grammatical study of language include the following:

- **Phonetics**, the study of the physical properties of speech sound production and perception, and delves into their acoustic and articulatory properties
- **Phonology**, the study of sounds as abstract elements in the speaker's mind that distinguish meaning (phonemes)
- **Morphology**, the study of morphemes, or the internal structures of words and how they can be modified
- **Syntax**, the study of how words combine to form grammatical phrases and sentences
- **Semantics**, the study of lexical and grammatical aspects of meaning
- **Pragmatics**, the study of how utterances are used in communicative acts, and the role played by situational context and non-linguistic knowledge in the transmission of meaning
- **Discourse analysis**, the analysis of language use in texts (spoken, written, or signed)
- **Stylistics**, the study of linguistic factors (rhetoric, diction, stress) that place a discourse in context
- **Semiotics**, the study of signs and sign processes (semiosis), indication, designation, likeness, analogy, metaphor, symbolism, signification, and communication

## **Discourse**

Discourse is language as social practice (Baynham, 1995) and is a multilayered concept. As a social practice, discourse embodies different ideologies through written and spoken texts. Discourse analysis can examine or expose these ideologies. Discourse influences genre, which is chosen in response to different situations and finally, at micro level, discourse influences language as text (spoken or written) at the phonological or lexico-grammatical level. Grammar and discourse are linked as parts of a system. A particular discourse becomes a language variety when it is used in this way for a particular purpose, and is referred to as a register. There may be certain lexical additions (new words) that are brought into play because of the expertise of the community of people within a certain domain of specialization. Registers and discourses therefore differentiate themselves through the use of vocabulary, and at times through the use of style too. People in the medical fraternity, for example, may use some medical terminology in their communication that is specialized to the field of medicine. This is often referred to as being part of the "medical discourse", and so on.

## **Lexicon**

The lexicon is a catalogue of words and terms that are stored in a speaker's mind. The lexicon consists of words and bound morphemes, which are parts of words that can't stand alone, like affixes. In some analyses, compound words and certain classes of idiomatic expressions and other collocations are also considered to be part of the lexicon. Dictionaries represent

attempts at listing, in alphabetical order, the lexicon of a given language; usually, however, bound morphemes are not included. Lexicography, closely linked with the domain of semantics, is the science of mapping the words into an encyclopedia or a dictionary. The creation and addition of new words (into the lexicon) is called coining or neologization, and the new words are called neologisms.

It is often believed that a speaker's capacity for language lies in the quantity of words stored in the lexicon. However, this is often considered a myth by linguists. The capacity for the use of language is considered by many linguists to lie primarily in the domain of grammar, and to be linked with competence, rather than with the growth of vocabulary. Even a very small lexicon is theoretically capable of producing an infinite number of sentences.

## **Style**

Stylistics also involves the study of written, signed, or spoken discourse through varying speech communities, genres, and editorial or narrative formats in the mass media. It involves the study and interpretation of texts for aspects of their linguistic and tonal style. Stylistic analysis entails the analysis of description of particular dialects and registers used by speech communities. Stylistic features include rhetoric, diction, stress, satire, irony, dialogue, and other forms of phonetic variations. Stylistic analysis can also include the study of language in canonical works of literature, popular fiction, news, advertisements, and other forms of communication in popular culture as well. It is usually seen as a variation in communication that changes from speaker to speaker and

community to community. In short, Stylistics is the interpretation of text.

In the 1960s, Jacques Derrida, for instance, further distinguished between speech and writing, by proposing that written language be studied as a linguistic medium of communication in itself. Palaeography is therefore the discipline that studies the evolution of written scripts (as signs and symbols) in language. The formal study of language also led to the growth of fields like psycholinguistics, which explores the representation and function of language in the mind; neurolinguistics, which studies language processing in the brain; biolinguistics, which studies the biology and evolution of language; and language acquisition, which investigates how children and adults acquire the knowledge of one or more languages.

## **Approaches**

### **Humanistic**

The fundamental principle of humanistic linguistics is that language is an invention created by people. A semiotic tradition of linguistic research considers language a sign system which arises from the interaction of meaning and form. The organisation of linguistic levels is considered computational. Linguistics is essentially seen as relating to social and cultural studies because different languages are shaped in social interaction by the speech community. Frameworks representing the humanistic view of language include structural linguistics, among others.



Structural analysis means dissecting each linguistic level: phonetic, morphological, syntactic, and discourse, to the smallest units. These are collected into inventories (e.g. phoneme, morpheme, lexical classes, phrase types) to study their interconnectedness within a hierarchy of structures and layers. Functional analysis adds to structural analysis the assignment of semantic and other functional roles that each unit may have. For example, a noun phrase may function as the subject or object of the sentence; or the agent or patient.

Functional linguistics, or functional grammar, is a branch of structural linguistics. In the humanistic reference, the terms structuralism and functionalism are related to their meaning in other human sciences. The difference between formal and functional structuralism lies in the way that the two approaches explain why languages have the properties they have. Functional explanation entails the idea that language is a tool for communication, or that communication is the primary function of language. Linguistic forms are consequently explained by an appeal to their functional value, or usefulness. Other structuralist approaches take the perspective that form follows from the inner mechanisms of the bilateral and multilayered language system.

## **Biological**

Approaches such as cognitive linguistics and generative grammar study linguistic cognition with a view towards uncovering the biological underpinnings of language. In Generative Grammar, these underpinning are understood as including innate domain-specific grammatical knowledge. Thus, one of the central concerns of the approach is to

discover what aspects of linguistic knowledge are innate and which are not.

Cognitive Linguistics, in contrast, rejects the notion of innate grammar, and studies how the human mind creates linguistic constructions from event schemas, and the impact of cognitive constraints and biases on human language. Similarly to neuro-linguistic programming, language is approached via the senses.

A closely related approach is evolutionary linguistics which includes the study of linguistic units as cultural replicators. It is possible to study how language replicates and adapts to the mind of the individual or the speech community. Construction grammar is a framework which applies the meme concept to the study of syntax.

The generative versus evolutionary approach are sometimes called formalism and functionalism, respectively. This reference is however different from the use of the terms in human sciences.

## **Methodology**

Linguistics is primarily descriptive. Linguists describe and explain features of language without making subjective judgments on whether a particular feature or usage is "good" or "bad". This is analogous to practice in other sciences: a zoologist studies the animal kingdom without making subjective judgments on whether a particular species is "better" or "worse" than another.

Prescription, on the other hand, is an attempt to promote particular linguistic usages over others, often favouring a particular dialect or "acrolect". This may have the aim of establishing a linguistic standard, which can aid communication over large geographical areas. It may also, however, be an attempt by speakers of one language or dialect to exert influence over speakers of other languages or dialects (see Linguistic imperialism). An extreme version of prescriptivism can be found among censors, who attempt to eradicate words and structures that they consider to be destructive to society. Prescription, however, may be practised appropriately in language instruction, like in ELT, where certain fundamental grammatical rules and lexical items need to be introduced to a second-language speaker who is attempting to acquire the language.

## **Sources**

Most contemporary linguists work under the assumption that spoken data and signed data are more fundamental than written data. This is because

- Speech appears to be universal to all human beings capable of producing and perceiving it, while there have been many cultures and speech communities that lack written communication;
- Features appear in speech which aren't always recorded in writing, including phonological rules, sound changes, and speech errors;
- All natural writing systems reflect a spoken language (or potentially a signed one), even with pictographic scripts like Dongba writing Naxi homophones with

the same pictogram, and text in writing systems used for two languages changing to fit the spoken language being recorded;

- Speech evolved before human beings invented writing;
- Individuals learn to speak and process spoken language more easily and earlier than they do with writing.

Nonetheless, linguists agree that the study of written language can be worthwhile and valuable. For research that relies on corpus linguistics and computational linguistics, written language is often much more convenient for processing large amounts of linguistic data. Large corpora of spoken language are difficult to create and hard to find, and are typically transcribed and written. In addition, linguists have turned to text-based discourse occurring in various formats of computer-mediated communication as a viable site for linguistic inquiry.

The study of writing systems themselves, graphemics, is, in any case, considered a branch of linguistics.

## **Analysis**

Before the 20th century, linguists analysed language on a diachronic plane, which was historical in focus. This meant that they would compare linguistic features and try to analyse language from the point of view of how it had changed between then and later. However, with Saussurean linguistics in the 20th century, the focus shifted to a more synchronic approach, where the study was more geared towards analysis and

comparison between different language variations, which existed at the same given point of time.

At another level, the syntagmatic plane of linguistic analysis entails the comparison between the way words are sequenced, within the syntax of a sentence. For example, the article "the" is followed by a noun, because of the syntagmatic relation between the words. The paradigmatic plane on the other hand, focuses on an analysis that is based on the paradigms or concepts that are embedded in a given text. In this case, words of the same type or class may be replaced in the text with each other to achieve the same conceptual understanding.

## **History**

The earliest activities in the description of language have been attributed to the 6th-century-BC Indian grammarian Pāṇini who wrote a formal description of the Sanskrit language in his *Aṣṭādhyāyī*. Today, modern-day theories on grammar employ many of the principles that were laid down back then.

## **Nomenclature**

Before the 20th century, the term *philology*, first attested in 1716, was commonly used to refer to the study of language, which was then predominantly historical in focus. Since Ferdinand de Saussure's insistence on the importance of synchronic analysis, however, this focus has shifted and the term *philology* is now generally used for the "study of a language's grammar, history, and literary tradition", especially

in the United States (where philology has never been very popularly considered as the "science of language").

Although the term "linguist" in the sense of "a student of language" dates from 1641, the term "linguistics" is first attested in 1847. It is now the usual term in English for the scientific study of language, though *linguistic science* is sometimes used.

Linguistics is a multi-disciplinary field of research that combines tools from natural sciences, social sciences, formal sciences, and the humanities. Many linguists, such as David Crystal, conceptualize the field as being primarily scientific. The term *linguist* applies to someone who studies language or is a researcher within the field, or to someone who uses the tools of the discipline to describe and analyse specific languages.

## **Early grammarians**

The formal study of language began in India with Pāṇini, the 6th century BC grammarian who formulated 3,959 rules of Sanskrit morphology. Pāṇini's systematic classification of the sounds of Sanskrit into consonants and vowels, and word classes, such as nouns and verbs, was the first known instance of its kind. In the Middle East, Sibawayh, a Persian, made a detailed description of Arabic in AD 760 in his monumental work, *Al-kitabfii an-naḥw* (الكتاب في النحو, *The Book on Grammar*), the first known author to distinguish between sounds and phonemes (sounds as units of a linguistic system). Western interest in the study of languages began somewhat later than in the East, but the grammarians of the classical

languages did not use the same methods or reach the same conclusions as their contemporaries in the Indic world. Early interest in language in the West was a part of philosophy, not of grammatical description. The first insights into semantic theory were made by Plato in his *Cratylus* dialogue, where he argues that words denote concepts that are eternal and exist in the world of ideas. This work is the first to use the word etymology to describe the history of a word's meaning. Around 280 BC, one of Alexander the Great's successors founded a university (see Musaeum) in Alexandria, where a school of philologists studied the ancient texts in and taught Greek to speakers of other languages. While this school was the first to use the word "grammar" in its modern sense, Plato had used the word in its original meaning as "téchnēgrammatikḗ" (ΤέχνηΓραμματική), the "art of writing", which is also the title of one of the most important works of the Alexandrine school by Dionysius Thrax. Throughout the Middle Ages, the study of language was subsumed under the topic of philology, the study of ancient languages and texts, practised by such educators as Roger Ascham, Wolfgang Ratke, and John Amos Comenius.

## **Comparative philology**

In the 18th century, the first use of the comparative method by William Jones sparked the rise of comparative linguistics. Bloomfield attributes "the first great scientific linguistic work of the world" to Jacob Grimm, who wrote *Deutsche Grammatik*. It was soon followed by other authors writing similar comparative studies on other language groups of Europe. The study of language was broadened from Indo-European to

language in general by Wilhelm von Humboldt, of whom Bloomfield asserts:

This study received its foundation at the hands of the Prussian statesman and scholar Wilhelm von Humboldt (1767–1835), especially in the first volume of his work on Kavi, the literary language of Java, entitled *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluß auf die geistige Entwicklung des Menschengeschlechts* (*On the Variety of the Structure of Human Language and its Influence upon the Mental Development of the Human Race*).

## **20th century developments**

There was a shift of focus from historical and comparative linguistics to synchronic analysis in early 20th century. Structural analysis was improved by Leonard Bloomfield, Louis Hjelmslev; and Zellig Harris who also developed methods of discourse analysis. Functional analysis was developed by the Prague linguistic circle and André Martinet. As sound recording devices became commonplace in the 1960s, dialectal recordings were made and archived, and the audio-lingual method provided a technological solution to foreign language learning. The 1960s also saw a new rise of comparative linguistics: the study of language universals in linguistic typology. Towards the end of the century the field of linguistics became divided into further areas of interest with the advent of language technology and digitalised corpora.



## **Areas of research**

Sociolinguistics is the study of how language is shaped by social factors. This sub-discipline focuses on the synchronic approach of linguistics, and looks at how a language in general, or a set of languages, display variation and varieties at a given point in time. The study of language variation and the different varieties of language through dialects, registers, and idiolects can be tackled through a study of style, as well as through analysis of discourse. Sociolinguists research both style and discourse in language, as well as the theoretical factors that are at play between language and society.

Developmental linguistics is the study of the development of linguistic ability in individuals, particularly the acquisition of language in childhood. Some of the questions that developmental linguistics looks into is how children acquire different languages, how adults can acquire a second language, and what the process of language acquisition is.

## **Neurolinguistics**

Neurolinguistics is the study of the structures in the human brain that underlie grammar and communication. 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 modelling. Amongst the structures of the brain involved in the mechanisms of neurolinguistics, the cerebellum which contains the highest numbers of neurons has a major role in terms of predictions required to produce language.

## **Applied linguistics**

Linguists are largely concerned with finding and describing the generalities and varieties both within particular languages and among all languages. Applied linguistics takes the results of those findings and "applies" them to other areas. Linguistic research is commonly applied to areas such as language education, lexicography, translation, language planning, which involves governmental policy implementation related to language use, and natural language processing. "Applied linguistics" has been argued to be something of a misnomer. Applied linguists actually focus on making sense of and engineering solutions for real-world linguistic problems, and not literally "applying" existing technical knowledge from linguistics. Moreover, they commonly apply technical knowledge from multiple sources, such as sociology (e.g., conversation analysis) and anthropology. (Constructed language fits under Applied linguistics.)

Today, computers are widely used in many areas of applied linguistics. Speech synthesis and speech recognition use phonetic and phonemic knowledge to provide voice interfaces to computers. Applications of computational linguistics in

machine translation, computer-assisted translation, and natural language processing are areas of applied linguistics that have come to the forefront. Their influence has had an effect on theories of syntax and semantics, as modelling syntactic and semantic theories on computers constraints.

Linguistic analysis is a sub-discipline of applied linguistics used by many governments to verify the claimed nationality of people seeking asylum who do not hold the necessary documentation to prove their claim. This often takes the form of an interview by personnel in an immigration department. Depending on the country, this interview is conducted either in the asylum seeker's native language through an interpreter or in an international *lingua franca* like English. Australia uses the former method, while Germany employs the latter; the Netherlands uses either method depending on the languages involved. Tape recordings of the interview then undergo language analysis, which can be done either by private contractors or within a department of the government. In this analysis, linguistic features of the asylum seeker are used by analysts to make a determination about the speaker's nationality. The reported findings of the linguistic analysis can play a critical role in the government's decision on the refugee status of the asylum seeker.

## **Language documentation**

Language documentation combines anthropological inquiry (into the history and culture of language) with linguistic inquiry, in order to describe languages and their grammars. Lexicography involves the documentation of words that form a vocabulary. Such a documentation of a linguistic vocabulary

from a particular language is usually compiled in a dictionary. Computational linguistics is concerned with the statistical or rule-based modeling of natural language from a computational perspective. Specific knowledge of language is applied by speakers during the act of translation and interpretation, as well as in language education – the teaching of a second or foreign language. Policy makers work with governments to implement new plans in education and teaching which are based on linguistic research.

Since the inception of the discipline of linguistics, linguists have been concerned with describing and analysing previously undocumented languages. Starting with Franz Boas in the early 1900s, this became the main focus of American linguistics until the rise of formal linguistics in the mid-20th century. This focus on language documentation was partly motivated by a concern to document the rapidly disappearing languages of indigenous peoples. The ethnographic dimension of the Boasian approach to language description played a role in the development of disciplines such as sociolinguistics, anthropological linguistics, and linguistic anthropology, which investigate the relations between language, culture, and society.

The emphasis on linguistic description and documentation has also gained prominence outside North America, with the documentation of rapidly dying indigenous languages becoming a focus in some university programmes in linguistics. Language description is a work-intensive endeavour, usually requiring years of field work in the language concerned, so as to equip the linguist to write a sufficiently accurate reference grammar. Further, the task of documentation requires the

linguist to collect a substantial corpus in the language in question, consisting of texts and recordings, both sound and video, which can be stored in an accessible format within open repositories, and used for further research.

## **Translation**

The sub-field of translation includes the translation of written and spoken texts across media, from digital to print and spoken. To translate literally means to transmute the meaning from one language into another. Translators are often employed by organizations such as travel agencies and governmental embassies to facilitate communication between two speakers who do not know each other's language. Translators are also employed to work within computational linguistics setups like Google Translate, which is an automated program to translate words and phrases between any two or more given languages. Translation is also conducted by publishing houses, which convert works of writing from one language to another in order to reach varied audiences. Academic translators specialize in or are familiar with various other disciplines such as technology, science, law, economics, etc.

## **Clinical linguistics**

Clinical linguistics is the application of linguistic theory to the field of speech-language pathology. Speech language pathologists work on corrective measures to treat communication and swallowing disorders.

Chaika (1990) showed that people with schizophrenia who display speech disorders like rhyming inappropriately have attentional dysfunction, as when a patient was shown a color chip and then asked to identify it, responded "looks like clay. Sounds like gray. Take you for a roll in the hay. Heyday, May Day." The color chip was actually clay-colored, so his first response was correct.'

However, most people suppress or ignore words which rhyme with what they've said unless they are deliberately producing a pun, poem or rap. Even then, the speaker shows connection between words chosen for rhyme and an overall meaning in discourse. People with schizophrenia with speech dysfunction show no such relation between rhyme and reason. Some even produce stretches of gibberish combined with recognizable words.

## **Computational linguistics**

Computational linguistics is the study of linguistic issues in a way that is "computationally responsible", i.e., taking careful note of computational consideration of algorithmic specification and computational complexity, so that the linguistic theories devised can be shown to exhibit certain desirable computational properties and their implementations. Computational linguists also work on computer language and software development.

## **Evolutionary linguistics**

Evolutionary linguistics is the study of the emergence of the language faculty through human evolution, and also the application of evolutionary theory to the study of cultural evolution among different languages. It is also a study of the dispersal of various languages across the globe, through movements among ancient communities. Evolutionary linguistics is a highly interdisciplinary field, including linguists, biologists, neuroscientists, psychologists, mathematicians, and others. By shifting the focus of investigation in linguistics to a comprehensive scheme that embraces the natural sciences, it seeks to yield a framework by which the fundamentals of language are understood.

## **Forensic linguistics**

Forensic linguistics is the application of linguistic analysis to forensics. Forensic analysis investigates the style, language, lexical use, and other linguistic and grammatical features used in the legal context to provide evidence in courts of law. Forensic linguists have also used their expertise in the framework of criminal cases.

## Chapter 2

# 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 Linguistics was 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'Association canadienne de linguistique appliquée (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/Cumann Teangeolaíochta Feidhmí (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.

## **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 bounded Turing 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 dichotomy 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 subdivisions 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 called conversational 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 language has 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 1950 Alan 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 Rogerian psychotherapist 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 Browning 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 1971 Terry 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 LUNAR was 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.

## Language assessment

**Language assessment** or **language testing** is a field of study under the umbrella of applied linguistics. Its main focus is the assessment of first, second or other language in the school, college, or university context; assessment of language use in the workplace; and assessment of language in the immigration, citizenship, and asylum contexts. The assessment may include listening, speaking, reading, writing, an integration of two or

more of these skills, or other constructs of language ability. Equal weight may be placed on knowledge (understanding how the language works theoretically) and proficiency (ability to use the language practically), or greater weight may be given to one aspect or the other.

## **History**

The history of language testing may have originated in the late nineteenth century testing of ESL at Cambridge and Oxford in England, but the earliest works in language assessment in the United States date back to the 1950s to the pioneering studies and test created by Robert Lado and David Harris. The earliest large scale assessments in the United States were referred to as the Michigan Tests, developed by the English Language Institute at the University of Michigan, now known as CaMLA, and the *Test of English as a Foreign Language* (TOEFL) developed by Educational Testing Service (ETS), Princeton, New Jersey.

The English Language Institute at the University of Michigan (CaMLA) was established in 1941 and was the first of its kind in the United States. Charles Fries, Director of ELI, and Robert Lado, Director of Testing at ELI, were determined to put foreign language teaching and testing on a "scientific" footing. The first test launched in 1946 was the *Lado Test of Aural Comprehension*. Approximately 10 years later, a full suite of tests had been assembled: "an English language test battery", which was administered to incoming foreign students at Michigan and other universities. Today this is known as the Michigan English Language Assessment Battery (MELAB). In

1953, the ELI also developed the ECPE (Examination for the Certificate of Proficiency in English) exam, under contract to the United States Information Agency, for use abroad.

TOEFL was launched in 1961 and was designed to assess the English language ability of students applying for admission to U.S. and Canadian colleges and universities. This test, which is used widely around the world, is still in use although it is now only available in the internet-based format (now called the TOEFL iBT).

Many tests from other companies, universities and agencies compete for this market: iTEP (International Test of English Proficiency), the Canadian English Language Proficiency Index Program (CELPIP) Test, the Pearson Language Test's *Pearson Test of English* (PTE), CaMLA assessments including the *Michigan English Language Assessment Battery* (MELAB) and Cambridge English Language Assessment, the British Council and the Australian IDP's *International English Language Testing System* (IELTS). In the United States, non-profit and other organizations such as the Center for Applied Linguistics, Washington, D.C. and Language Testing International, White Plains, NY have developed language tests that are used by many public and private agencies. Many universities too, like the University of California, Los Angeles, Teachers College, Columbia University, and the University of Illinois, Urbana-Champaign, have developed English (and other) language tests to assess the abilities of their students and teaching assistants. These language assessments are generally known as proficiency or achievement assessments. Other modern English language tests developed include *The General English Proficiency Test* (GEPT) in Taiwan, the *College English Test* in

China, and the *STEP Eiken* in Japan. New technology has also made a presence in the field: Versant's English and Dutch assessments use phone technology to record the speaking and automated scoring of their speaking tests, and the ETS is currently experimenting with automated scoring of their writing tests.

## **Organizations**

The International Language Testing Association (ILTA) is one of the many organizations that organizes conferences, workshops, and a public forum for the discussion of important matters. ILTA's major annual conference is the Language Testing Research Colloquium. ILTA's Lifetime Achievement Award winners include: Carol Chapelle (USA), Alan Davies (UK), Lyle Bachman (USA), Bernard Spolsky (Israel), John Clark (USA), Charles Alderson (UK) and Elana Shohamy (Israel).

Educational Testing Service, Princeton, New Jersey, the home of the TOEFL, offers an annual outstanding Doctoral Dissertation Award in Second or Foreign Language and the University of Cambridge, UK, also offers an annual outstanding master's degree Award in second language testing. In Europe, there are two organizations: the Association of Language Testers of Europe (ALTE) and the European Association for Language Testing and Assessment (EALTA). All of these associations have developed Codes of Ethics and Practice that all language assessment professionals are expected to adhere to.



## **Annual conferences**

There are many annual conferences on general or specific topics. Among the most important conferences is ILTA's official conference: the **Language Testing Research Colloquium** (LTRC), which has been held every year since 1978. In the last few years, it has been held in different parts of the world: Temecula, California, USA (2004); Ottawa, Canada (2005); Melbourne, Australia (2006); Barcelona, Spain (2007); Hangzhou, China (2008), and Denver, Colorado (2009), Cambridge, UK (2010), Ann Arbor, Michigan (2011), Princeton, New Jersey (2012), and Seoul, South Korea (2013).

ALTE's international conferences are held in different cities in Europe: Barcelona, Spain (2002); Berlin, Germany (2005); Cambridge, UK (2008); Krakow, Poland (2011); Paris, France (2014); Bologna, Italy (2017) ) with regional conferences in Perugia, Prague, Budapest, Sofia, and Lisbon. Similarly, there are regional meetings in China, Japan, Korea, and Taiwan. International conference themes have included supporting the European Year of Languages (2001), the impact of multilingualism (2005), the wider social and educational impact of assessment (2008) and the role of language frameworks (2011). Selected conference papers have been published through the Studies in Language Testing (SiLT) volumes.

# **Language assessments in aviation**

The aviation personnel is required to be regularly tested on aviation language proficiency. The testing is required on the international basis by ICAO Doc 9835. Within the EASA region, the aviation language assessments are required by EC 1178/2011, Part FCL, FCL.055.

## **Endorsements**

Each flight crew license is endorsed by the respective endorsement specifying the holder's language proficiency. The level acceptable for the operational use in the aviation is 4 thru 6. The validity of the respective level varies with the region.

## **Organizations**

The organizations authorized to conduct the language assessments on behalf of national aviation authorities are so-called "Language assessment bodies" or "Testing service providers". Each and every Language assessment body is issued with the Certificate of approval with its authorizations. Aero Language and Myflower College are one of several organizations in Europe authorized to conduct the language assessments for pilots and air traffic controllers.

## Publications

There are two premier journals in the field: *Language Assessment Quarterly* (published by Routledge/Taylor & Francis) and *Language Testing* (published by Sage Publications) that publishes major findings from researchers. Both these journals are indexed in Thompson's SSCI list. Other journals that publish articles from the field include *Applied Linguistics*, *Language Learning*, *TESOL Quarterly*, *Assessing Writing*, and *System*. Some of these journals have special issue volumes on Ethics in language assessment, structural equation modeling, language assessment in Asia, Classroom assessment, etc. and commentaries, brief reports, and book and test reviews.

The field has exploded in the last twenty years in terms of textbooks and research publications. The most popular books include: Lyle Bachman's *Fundamental considerations in language testing*, and *Statistical Analyses for Language Assessment*, Lyle Bachman and Adrian Palmer's *Language Testing in Practice* and *Language Assessment in Practice*, Glenn Fulcher and Fred Davidson's *'Language Testing and Assessment: An Advanced Resource Book'*, Charles Alderson's *'Assessing Reading*, John Read's *Assessing Vocabulary*, James Purpura's *Assessing Grammar*, Gary Buck's *Assessing Listening*, Sara Weigle's *Assessing Writing*, Glenn Fulcher's *'Practical Language Testing'* and *'Testing Second Language Speaking'*. Edited volumes include: Alister Cumming's *Validation in Language Testing*, Antony John Kunnan's *Validation in Language Assessment*, and *Fairness in Language Assessment*,

*and the 'Routledge Handbook of Language Testing', edited by Glenn Fulcher and Fred Davidson.*

The most popular book series are Michael Milanovic, Cyril Weir, and Lynda Taylor's *Studies in Language Testing (SiLT)* series, and Lyle Bachman and Charles Alderson's *Cambridge Language Assessment Series*.

## **Courses**

Language assessment or language testing courses are taught as required or elective courses in many graduate and doctoral programs, particularly in the subjects of applied linguistics, English for Speakers of Other Languages, English as a second or foreign language, or educational linguistics. These programs are known as MA or PhD programs in Applied Linguistics, Educational Linguistics, TESOL, TEFL, or TESL. The focus of most courses is on test development, psychometric qualities of tests, validity, reliability and fairness of tests, and classical true score measurement theory. Additional courses focus on item response theory, factor analysis, structural equation modeling, G theory, latent growth modeling, qualitative analysis of test performance data such as conversation and discourse analysis, and politics and language policy issues.

Universities that have regular courses and programs that focus on language assessment at the PhD level include Iowa State University, University of Illinois, Urbana-Champaign, University of Hawai'i, Manoa, Teachers College, Columbia University, Penn State University, Georgia State University, Northern Arizona University, McGill University, University of

Toronto, Lancaster University (UK), University of Leicester, University of Bristol, University of Cambridge, University of Bedfordshire, and Guangdong University of Foreign Studies (China); at the MA level include Lancaster University, University of Leicester, California State Universities at Fullerton, Los Angeles, Long Beach, San Jose, and San Francisco.

## Other Test Types

- Test of English for International Communication (TOEIC)

Accepted and trusted by 14,000+ organizations in more than 160 countries, the TOEIC® tests assess your English-language proficiency across all four language skills needed to succeed in the global workplace — listening, reading, speaking and writing. With your TOEIC score, you can: get a fair, accurate evaluation of your ability to communicate in English; show potential employers your full range of communication skills; differentiate yourself from the competition; expand your job opportunities.

## 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

# Language education

**Language education** – the process and practice of teaching a second or foreign language – is primarily a branch of applied linguistics, but can be an interdisciplinary field. There are four main learning categories for language education: communicative competencies, proficiencies, cross-cultural experiences, and multiple literacies.

# **Need**

Increasing globalization has created a great need for people in the workforce who can communicate in multiple languages. Common languages are used in areas such as trade, tourism, international relations, technology, media, and science. Many countries such as Korea (Kim Yeong-seo, 2009), Japan (Kubota, 1998) and China (Kirkpatrick & Zhichang, 2002) frame education policies to teach at least one foreign language at the primary and secondary school levels. However, some countries such as India, Singapore, Malaysia, Pakistan, and the Philippines use a second official language in their governments. According to GAO (2010), China has recently been putting enormous importance on foreign language learning, especially the English language.

# **History**

## **Ancient to medieval period**

The need to learn foreign languages is as old as human history itself. In the Ancient Near East, Akkadian was the language of diplomacy, as in the Amarna letters. For many centuries, Latin was the dominant language of education, commerce, religion, and government in much of Europe, but it was displaced for many purposes by French, Italian, and English by the end of the 16th century. John Amos Comenius was one of many people who tried to reverse this trend. He wrote a complete

course for learning Latin, covering the entire school curriculum, culminating in his *Opera Didactica Omnia*, 1657.

In this work, Comenius also outlined his theory of language acquisition. He is one of the first theorists to write systematically about how languages are learned and about methods for teaching languages. He held that language acquisition must be allied with sensation and experience. Teaching must be oral. The schoolroom should have models of things, or else pictures of them. He published the world's first illustrated children's book, *Orbissensualium pictus*. The study of Latin gradually diminished from the study of a living language to a mere subject in the school curriculum. This decline demanded a new justification for its study. It was then claimed that the study of Latin developed intellectual ability, and the study of Latin grammar became an end in and of itself.

"Grammar schools" from the 16th to 18th centuries focused on teaching the grammatical aspects of Classical Latin. Advanced students continued grammar study with the addition of rhetoric.

## **18th century**

The study of modern languages did not become part of the curriculum of European schools until the 18th century. Based on the purely academic study of Latin, students of modern languages did much of the same exercises, studying grammatical rules and translating abstract sentences. Oral work was minimal, and students were instead required to memorize grammatical rules and apply these to decode written

texts in the target language. This tradition-inspired method became known as the grammar-translation method.

## **19th and 20th centuries**

Innovation in foreign language teaching began in the 19th century and became very rapid in the 20th century. It led to a number of different and sometimes conflicting methods, each claiming to be a major improvement over the previous or contemporary methods. The earliest applied linguists included Jean Manes ca, Heinrich Gottfried Ollendorff (1803–1865), Henry Sweet (1845–1912), Otto Jespersen (1860–1943), and Harold Palmer (1877–1949). They worked on setting language teaching principles and approaches based on linguistic and psychological theories, but they left many of the specific practical details for others to devise.

The history of foreign-language education in the 20th century and the methods of teaching (such as those related below) might appear to be a history of failure. Very few students in U.S. universities who have a foreign language as a major attain "minimum professional proficiency". Even the "reading knowledge" required for a PhD degree is comparable only to what second-year language students read, and only very few researchers who are native English speakers can read and assess information written in languages other than English. Even a number of famous linguists are monolingual.

However, anecdotal evidence for successful second or foreign language learning is easy to find, leading to a discrepancy between these cases and the failure of most language programs. This tends to make the research of second language

acquisition emotionally charged. Older methods and approaches such as the grammar translation method and the direct method are dismissed and even ridiculed, as newer methods and approaches are invented and promoted as the only and complete solution to the problem of the high failure rates of foreign language students.

Most books on language teaching list the various methods that have been used in the past, often ending with the author's new method. These new methods are usually presented as coming only from the author's mind, as the authors generally give no credence to what was done before and do not explain how it relates to the new method. For example, descriptive linguists seem to claim unhesitatingly that there were no scientifically based language teaching methods before their work (which led to the audio-lingual method developed for the U.S. Army in World War II). However, there is significant evidence to the contrary. It is also often inferred or even stated that older methods were completely ineffective or have died out completely, though in reality even the oldest methods are still in use (e.g. the Berlitz version of the direct method). Proponents of new methods have been so sure that their ideas are so new and so correct that they could not conceive that the older ones have enough validity to cause controversy. This was in turn caused by emphasis on new scientific advances, which has tended to blind researchers to precedents in older work.(p. 5)

There have been two major branches in the field of language learning, the empirical and theoretical, and these have almost completely separate histories, with each gaining ground over the other at one time or another. Examples of researchers on

the empiricist side are Jespersen, Palmer, and Leonard Bloomfield, who promote mimicry and memorization with pattern drills. These methods follow from the basic empiricist position that language acquisition results from habits formed by conditioning and drilling. In its most extreme form, language learning is seen as much the same as any other learning in any other species, human language being essentially the same as communication behaviors seen in other species.

On the theoretical side are, for example, Francois Gouin, M.D. Berlitz, and Emile B. De Sauzé, whose rationalist theories of language acquisition dovetail with linguistic work done by Noam Chomsky and others. These have led to a wider variety of teaching methods, ranging from the grammar-translation method and Gouin's "series method" to the direct methods of Berlitz and De Sauzé. With these methods, students generate original and meaningful sentences to gain a functional knowledge of the rules of grammar. This follows from the rationalist position that man is born to think and that language use is a uniquely human trait impossible in other species. Given that human languages share many common traits, the idea is that humans share a universal grammar which is built into our brain structure. This allows us to create sentences that we have never heard before but that can still be immediately understood by anyone who understands the specific language being spoken. The rivalry between the two camps is intense, with little communication or cooperation between them.

## **21st century**

Over time, language education has developed in schools and has become a part of the education curriculum around the world. In some countries, such as the United States, language education (also referred to as World Languages) has become a core subject along with main subjects such as English, Maths and Science.

In some countries, such as Australia, it is so common nowadays for a foreign language to be taught in schools that the subject of language education is referred to LOTE or Language Other Than English. In the majority of English-speaking education centers, French, Spanish, and German are the most popular languages to study and learn. English as a Second Language (ESL) is also available for students whose first language is not English and they are unable to speak it to the required standard.

## **Teaching foreign language in classrooms**

Language education may take place as a general school subject or in a specialized **language school**. There are many methods of teaching languages. Some have fallen into relative obscurity and others are widely used; still others have a small following, but offer useful insights.



While sometimes confused, the terms "approach", "method" and "technique" are hierarchical concepts.

An **approach** is a set of assumptions about the nature of language and language learning, but does not involve procedure or provide any details about how such assumptions should be implemented into the classroom setting. Such can be related to second language acquisition theory.

There are three principal "approaches":

- The structural view treats language as a system of structurally related elements to code meaning (e.g. grammar).
- The functional view sees language as a vehicle to express or accomplish a certain function, such as requesting something.
- The interactive view sees language as a vehicle for the creation and maintenance of social relations, focusing on patterns of moves, acts, negotiation and interaction found in conversational exchanges. This approach has been fairly dominant since the 1980s.

A **method** is a plan for presenting the language material to be learned, and should be based upon a selected approach. In order for an approach to be translated into a method, an instructional system must be designed considering the objectives of the teaching/learning, how the content is to be selected and organized, the types of tasks to be performed, the roles of students, and the roles of teachers.

- Examples of structural methods are grammar translation and the audio-lingual method.

- Examples of functional methods include the oral approach / situational language teaching.
- Examples of interactive methods include the direct method, the series method, communicative language teaching, language immersion, the Silent Way, Suggestopedia, the Natural Approach, Tandem Language Learning, Total Physical Response, Teaching Proficiency through Reading and Storytelling and Dogme language teaching.

A **technique** (or strategy) is a very specific, concrete stratagem or trick designed to accomplish an immediate objective. Such are derived from the controlling method, and less directly, from the approach.

## **Online and self-study courses**

Hundreds of languages are available for self-study, from scores of publishers, for a range of costs, using a variety of methods. The course itself acts as a teacher and has to choose a methodology, just as classroom teachers do.

## **Audio recordings and books**

Audio **recordings** use native speakers, and one strength is helping learners improve their accent. Some recordings have pauses for the learner to speak. Others are continuous so the learner speaks along with the recorded voice, similar to learning a song.

Audio recordings for self-study use many of the methods used in classroom teaching, and have been produced on records, tapes, CDs, DVDs and websites.

Most audio recordings teach words in the target language by using explanations in the learner's own language. An alternative is to use sound effects to show meaning of words in the target language. The only language in such recordings is the target language, and they are comprehensible regardless of the learner's native language.

Language **books** have been published for centuries, teaching vocabulary and grammar. The simplest books are phrasebooks to give useful short phrases for travelers, cooks, receptionists, or others who need specific vocabulary. More complete books include more vocabulary, grammar, exercises, translation, and writing practice.

Also, various other "language learning tools" have been entering the market in recent years.

## **Internet and software**

**Software** can interact with learners in ways that books and audio cannot:

- Some software records the learner, analyzes the pronunciation, and gives feedback.
- Software can present additional exercises in areas where a particular learner has difficulty, until the concepts are mastered.

- Software can pronounce words in the target language and show their meaning by using pictures instead of oral explanations. The only language in such software is the target language. It is comprehensible regardless of the learner's native language.

**Websites** provide various services geared toward language education. Some sites are designed specifically for learning languages:

- Some software runs on the web itself, with the advantage of avoiding downloads, and the disadvantage of requiring an internet connection.
- Some publishers use the web to distribute audio, texts and software, for use offline. For example, various travel guides, for example Lonely Planet, offer software supporting language education.
- Some websites offer learning activities such as quizzes or puzzles to practice language concepts.
- Language exchange sites connect users with complementary language skills, such as a native Spanish speaker who wants to learn English with a native English speaker who wants to learn Spanish. Language exchange websites essentially treat *knowledge of a language* as a commodity, and provide a market like environment for the commodity to be exchanged. Users typically contact each other via chat, VoIP, or email. Language exchanges have also been viewed as a helpful tool to aid language learning at language schools. Language exchanges tend to benefit oral proficiency, fluency, colloquial vocabulary acquisition, and vernacular usage, rather

than formal grammar or writing skills. Across Australasia, 'Education Perfect' – an online learning site- is frequently used as it enables teachers to monitor students' progress as students gain a "point" for every new word remembered. There is an annual international Education Perfect languages contest held in May.

Many other websites are helpful for learning languages, even though they are designed, maintained and marketed for other purposes:

- All countries have websites in their own languages, which learners elsewhere can use as primary material for study: news, fiction, videos, songs, etc. In a study conducted by the Center for Applied Linguistics, it was noted that the use of technology and media has begun to play a heavy role in facilitating language learning in the classroom. With the help of the internet, students are readily exposed to foreign media (music videos, television shows, films) and as a result, teachers are taking heed of the internet's influence and are searching for ways to combine this exposure into their classroom teaching.
- Translation sites let learners find the meaning of foreign text or create foreign translations of text from their native language.
- Speech synthesis or text to speech (TTS) sites and software let learners hear pronunciation of arbitrary written text, with pronunciation similar to a native speaker.

- Course development and learning management systems such as Moodle are used by teachers, including language teachers.
- Web conferencing tools can bring remote learners together; e.g. Elluminate Live.
- Players of computer games can practice a target language when interacting in massively multiplayer online games and virtual worlds. In 2005, the virtual world Second Life started to be used for foreign language tuition, sometimes with entire businesses being developed. In addition, Spain's language and cultural institute Instituto Cervantes has an "island" on Second Life.

Some Internet content is free, often from government and nonprofit sites such as BBC Online, Book2, Foreign Service Institute, with no or minimal ads. Some are ad-supported, such as newspapers and YouTube. Some require a payment.

## **Learning strategies**

Language learning strategies have attracted increasing focus as a way of understanding the process of language acquisition.

### **Listening as a way to learn**

Clearly listening is used to learn, but not all language learners employ it consciously. Listening to understand is one level of listening but focused listening is not something that most learners employ as a strategy. Focused listening is a strategy in listening that helps students listen attentively with no

distractions. Focused listening is very important when learning a foreign language as the slightest accent on a word can change the meaning completely.

## **Reading as a way to learn**

Many people read to understand but the strategy of reading text to learn grammar and discourse styles can also be employed.

## **Speaking as a way to learn**

Alongside listening and reading exercises, practicing conversation skills is an important aspect of language acquisition. Language learners can gain experience in speaking foreign languages through in-person language classes, language meet-ups, university language exchange programs, joining online language learning communities (e.g. Conversation Exchange and Tandem), and traveling to a country where the language is spoken.

## **Learning vocabulary**

Translation and rote memorization have been the two strategies that have been employed traditionally. There are other strategies that also can be used such as guessing, based on looking for contextual clues, spaced repetition with a use of various apps, games and tools (e.g. DuoLingo, LingoMonkey and Vocabulary Stickers). Knowledge about how the brain works can be utilized in creating strategies for how to remember words.

## **Learning Esperanto**

Esperanto, the most widely used international auxiliary language, was founded by L. L. Zamenhof, a Polish-Jewish ophthalmologist, in 1887, aimed to eliminate language barriers in the international contacts. Esperanto is an artificial language created on the basis of the Indo-European languages, absorbing the reasonable factors of commonality of the Germanic languages. Esperanto is completely consistent in its speech and writing. The stress of every word is fixed on the penultimate syllable. By learning twenty-eight letters and mastering the phonetic rules, one can read and write any words. With further simplification and standardization, Esperanto becomes more easily mastered than other languages. Ease of learning helps one build the confidence and learning Esperanto, as a learning strategy, constitutes a good introduction to foreign language study.

## **Teaching strategies**

### **Blended learning**

Blended learning combines face-to-face teaching with distance education, frequently electronic, either computer-based or web-based. It has been a major growth point in the ELT (English Language Teaching) industry over the last ten years.

Some people, though, use the phrase 'Blended Learning' to refer to learning taking place while the focus is on other activities. For example, playing a card game that requires



calling for cards may allow blended learning of numbers (1 to 10).

## **Skill teaching**

When talking about language skills, the four basic ones are: listening, speaking, reading and writing. However, other, more socially based skills have been identified more recently such as summarizing, describing, narrating etc. In addition, more general learning skills such as study skills and knowing how one learns have been applied to language classrooms.

In the 1970s and 1980s, the four basic skills were generally taught in isolation in a very rigid order, such as listening before speaking. However, since then, it has been recognized that we generally use more than one skill at a time, leading to more integrated exercises. Speaking is a skill that often is underrepresented in the traditional classroom. This is due to the fact that it is considered harder to teach and test. There are numerous texts on teaching and testing writing but relatively few on speaking.

More recent textbooks stress the importance of students working with other students in pairs and groups, sometimes the entire class. Pair and group work give opportunities for more students to participate more actively. However, supervision of pairs and groups is important to make sure everyone participates as equally as possible. Such activities also provide opportunities for peer teaching, where weaker learners can find support from stronger classmates.

## **Sandwich technique**

In foreign language teaching, the **sandwich technique** is the oral insertion of an idiomatic translation in the mother tongue between an unknown phrase in the learned language and its repetition, in order to convey meaning as rapidly and completely as possible. The mother tongue equivalent can be given almost as an aside, with a slight break in the flow of speech to mark it as an intruder.

When modeling a dialogue sentence for students to repeat, the teacher not only gives an oral mother tongue equivalent for unknown words or phrases, but repeats the foreign language phrase before students imitate it: L2 => L1 => L2. For example, a German teacher of English might engage in the following exchange with the students:

- Teacher: "Let me try – lass michversuchen – let me try."
- Students: "Let me try."

## **Mother tongue mirroring**

Mother tongue mirroring is the adaptation of the time-honoured technique of literal translation or word-for word translation for pedagogical purposes. The aim is to make foreign constructions salient and transparent to learners and, in many cases, spare them the technical jargon of grammatical analysis. It differs from literal translation and interlinear text as used in the past since it takes the progress learners have made into account and only focuses upon a specific structure

at a time. As a didactic device, it can only be used to the extent that it remains intelligible to the learner, unless it is combined with a normal idiomatic translation. This technique is seldom referred to or used these days.

## **Back-chaining**

Back-chaining is a technique used in teaching oral language skills, especially with polysyllabic or difficult words. The teacher pronounces the last syllable, the student repeats, and then the teacher continues, working backwards from the end of the word to the beginning.

For example, to teach the name 'Mussorgsky' a teacher will pronounce the last syllable: *-sky*, and have the student repeat it. Then the teacher will repeat it with *-sorg-* attached before: *-sorg-sky*, and all that remains is the first syllable: *Mus-sorg-sky*.

## **Code Switching**

Code switching is a special linguistic phenomenon that the speaker consciously alternates two or more languages according to different time, places, contents, objects and other factors. Code switching shows its functions while one is in the environment that mother tongue are not playing a dominant role in students' life and study, such as the children in the bilingual family or in the immigrant family. That is to say, the capability of using code switching, relating to the transformation of phonetics, words, language structure, expression mode, thinking mode, cultural differences and so

on, is needed to be guided and developed in the daily communication environment. Most people learn foreign language in the circumstance filled with the using of their native language so that their ability of code switching cannot be stimulated, and thus the efficiency of foreign language acquisition would decrease. Therefore, as a teaching strategy, code switching is used to help students better gain conceptual competences and to provide rich semantic context for them to understand some specific vocabularies.

## **By region**

Practices in language education may vary by region however the underlying understandings which drive it are fundamentally similar. Rote repetition, drilling, memorisation and grammar conjugating are used the world over. Sometimes there are different preferences teaching methods by region. Language immersion is popular in some European countries, but is not used very much in the United States, in Asia or in Australia.

## **By different life stage**

### **Early childhood education**

Early childhood is the fastest and most critical period for one to master language in their life. Children's language communication is transformed from non-verbal communication to verbal communication from ages of one to five. Their

mastery of language is largely acquired naturally by living in a verbal communication environment. As long as we are good at guiding and creating opportunities for children, children's language ability is easy to be developed and cultivated.

## **Compulsory education**

Compulsory education, for most people, is the period that they have access to a second or foreign language for the first time. In this period, the most professional foreign language education and academic atmosphere are provided to the students. They can get help and motivation from teachers and be activated by the peers at any time. One would be able to undergo a lot of specialized learning in order to truly master a great number of rules of vocabulary, grammar and verbal communication.

## **Adult education**

Learning a foreign language during adulthood means one is pursuing a higher value of themselves by obtaining a new skill. At this stage, individuals have already developed the ability to supervise themselves learning a language. However, at the same time, the pressure is also an obstacle for adults.

## **Elderly education**

Compared to other life stages, this period is the hardest to learn a new language due to gradual brain deterioration and memory loss. Notwithstanding its difficulty, language

education for seniors can slow this brain degeneration and active ageing.

## **Language study holidays**

An increasing number of people are now combining holidays with language study in the native country. This enables the student to experience the target culture by meeting local people. Such a holiday often combines formal lessons, cultural excursions, leisure activities, and a homestay, perhaps with time to travel in the country afterwards. Language study holidays are popular across Europe (Malta & UK being the most popular because almost everyone speaks English as a first language) and Asia due to the ease of transportation and variety of nearby countries. These holidays have become increasingly more popular in Central and South America in such countries as Guatemala, Ecuador and Peru. As a consequence of this increasing popularity, several international language education agencies have flourished in recent years. Though education systems around the world invest enormous sums of money into language teaching the outcomes in terms of getting students to actually speak the language(s) they are learning outside the classroom are often unclear.

With the increasing prevalence of international business transactions, it is now important to have multiple languages at one's disposal. This is also evident in businesses outsourcing their departments to Eastern Europe.

# **Minority language education**

## **Minority language education policy**

The principal policy arguments in favor of promoting minority language education are the need for multilingual workforces, intellectual and cultural benefits and greater inclusion in global information society. Access to education in a minority language is also seen as a human right as granted by the European Convention on Human Rights and Fundamental Freedoms, the European Charter for Regional or Minority Languages and the UN Human Rights Committee. Bilingual Education has been implemented in many countries including the United States, in order to promote both the use and appreciation of the minority language, as well as the majority language concerned.

## **Materials and e-learning for minority language education**

Suitable resources for teaching and learning minority languages can be difficult to find and access, which has led to calls for the increased development of materials for minority language teaching. The internet offers opportunities to access a wider range of texts, audios and videos. Language learning 2.0 (the use of web 2.0 tools for language education) offers opportunities for material development for lesser-taught languages and to bring together geographically dispersed teachers and learners.

# Acronyms and abbreviations

- **ALL:** Apprenticeship Language Learning
- **CALL:** computer-assisted language learning
- **CLIL:** content and language integrated learning
- **CELI:** Certificato di Conoscenza della Lingua Italiana
- **CLL:** community language learning
- **DELE:** Diploma de Español como Lengua Extranjera
- **DELF:** diplôme d'études en langue française
- **EFL:** English as a foreign language
- **EAL/D:** English as an additional language or dialect
- **EAP:** English for academic purposes
- **ELL:** English language learning
- **ELT:** English language teaching
- **ESL:** English as a second language
- **ESP:** English for specific purposes      English for specific purposes
- **FLL:** foreign language learning
- **FLT:** foreign language teaching
- **HLL:** heritage language learning
- **IATEFL:** International Association of Teachers of English as a Foreign Language  
International Association of Teachers of English as a Foreign Language
- **L1:** first language, native language, mother tongue
- **L2:** second language (or any additional language)
- **LDL:** Lernend durch Lehren (German for learning by teaching)
- **LOTE:** Languages Other Than English
- **SLA:** second language acquisition



- **TELL**: technology-enhanced language learning
- **TEFL**: teaching English as a foreign language
- **TEFLA**: teaching English as a foreign language to adults
- **TESOL**: teaching English to speakers of other languages
- **TEYL**: teaching English to young learners
- **TPR**: Total Physical Response
- **TPRS**: Teaching Proficiency through Reading and Storytelling
- UNiCert is a European language education system of many universities based on the Common European Framework of Reference for Languages.
- See also English language learning and teaching for information on language teaching acronyms and abbreviations which are specific to English

## **Linguistic anthropology**

**Linguistic anthropology** is the interdisciplinary study of how language influences social life. It is a branch of anthropology that originated from the endeavor to document endangered languages, and has grown over the past century to encompass most aspects of language structure and use.

Linguistic anthropology explores how language shapes communication, forms social identity and group membership, organizes large-scale cultural beliefs and ideologies, and develops a common cultural representation of natural and social worlds.

# **Historical development**

Linguistic anthropology emerged from the development of three distinct paradigms. These paradigms set the ways of approaching linguistic anthropology: the first, now known as "anthropological linguistics", focuses on the documentation of languages; the second, known as "linguistic anthropology", engages in theoretical studies of language use; the third, developed over the past two or three decades, studies issues from other sub-fields of anthropology with linguistic tactics. Though they developed sequentially, all three paradigms are still practiced today.

## **First paradigm: Anthropological linguistics**

- Main article: Anthropological linguistics

The first paradigm is known as anthropological linguistics. The field is devoted to themes unique to the sub-discipline: documentation of languages that were then seen as doomed to extinction, with special focus on the languages of native North American tribes. It is also the paradigm most focused on linguistics. The themes include:

- Grammatical description,
- Typological classification and
- Linguistic relativity

## **Second paradigm: Linguistic anthropology**

The second paradigm can be marked by the switch from *anthropological linguistics* to *linguistic anthropology*, signalling a more anthropological focus on the study. This term was preferred by Dell Hymes, who was also responsible, with John Gumperz, for the idea of ethnography of communication. The term *linguistic anthropology* reflected Hymes' vision for the future, where language would be studied in the context of the situation, and relative to the community speaking it. This new era would involve many new technological developments, such as mechanical recording.

Hymes had many revolutionary contributions to linguistic anthropology, the first of which was a new unit of analysis. Unlike the first paradigm, which focused on linguistic tools like measuring of phonemes and morphemes, the second paradigm's unit of analysis was the "speech event". A speech event is an event defined by speech occurring during it (ex. a lecture, debate). This is different from a speech situation, where speech could possibly occur (ex. dinner). Hymes also pioneered a linguistic anthropological approach to ethnopoetics. Hymes had hoped that this paradigm would link linguistic anthropology more to anthropology. However, Hymes' ambition backfired as the second paradigm marked a distancing of the sub-discipline from the rest of anthropology.

## **Third paradigm: Anthropological issues studied via linguistic methods and data**

The third paradigm, which began in the late 1980s, refocused on anthropology by providing a linguistic approach to anthropological issues. Rather than focusing on exploring language, third paradigm anthropologists focus on studying culture with linguistic tools. Themes include:

- investigations of personal and social identities
- shared ideologies
- construction of narrative interactions among individuals

Furthermore, like how the second paradigm made use of new technology in its studies, the third paradigm heavily includes use of video documentation to support research.

## **Areas of interest**

Contemporary linguistic anthropology continues research in all three of the paradigms described above: documentation of languages, study of language through context, and study of identity through linguistic means. The third paradigm, the study of anthropological issues, is a particularly rich area of study for current linguistic anthropologists.

## **Identity and inter subjectivity**

A great deal of work in linguistic anthropology investigates questions of sociocultural identity linguistically and discursively. Linguistic anthropologist Don Kulick has done so in relation to identity, for example, in a series of settings, first in a village called Gapun in northern Papua New Guinea. He explored how the use of two languages with and around children in Gapun village: the traditional language (Taiap), not spoken anywhere but in their own village and thus primordially "indexical" of Gapuner identity, and TokPisin, the widely circulating official language of New Guinea. ("indexical" points to meanings beyond the immediate context.) To speak the Taiap language is associated with one identity: not only local but "Backward" and also an identity based on the display of \*hed\* (personal autonomy). To speak TokPisin is to index a modern, Christian (Catholic) identity, based not on \*hed\* but on \*save\*, an identity linked with the will and the skill to cooperate. In later work, Kulick demonstrates that certain loud speech performances in Brazil called \*um escândalo\*, Brazilian *travesti* (roughly, 'transvestite') sex workers shame clients. The *travesti* community, the argument goes, ends up at least making a powerful attempt to transcend the shame the larger Brazilian public might try to foist off on them, again by loud public discourse and other modes of performance.

In addition, scholars such as Émile Benveniste, Mary Bucholtz and Kira Hall Benjamin Lee, Paul Kockelman, and Stanton Wortham (among many others) have contributed to understandings of identity as "intersubjectivity" by examining the ways it is discursively constructed.

## **Socialization**

In a series of studies, linguistic anthropologists Elinor Ochs and Bambi Schieffelin addressed the anthropological topic of socialization (the process by which infants, children, and foreigners become members of a community, learning to participate in its culture), using linguistic and other ethnographic methods. They discovered that the processes of enculturation and socialization do not occur apart from the process of language acquisition, but that children acquire language and culture together in what amounts to an integrated process. Ochs and Schieffelin demonstrated that baby talk is not universal, that the direction of adaptation (whether the child is made to adapt to the ongoing situation of speech around it or vice versa) was a variable that correlated, for example, with the direction it was held vis-à-vis a caregiver's body. In many societies caregivers hold a child facing outward so as to orient it to a network of kin whom it must learn to recognize early in life.

Ochs and Schieffelin demonstrated that members of all societies socialize children both *to* and *through* the use of language. Ochs and Schieffelin uncovered how, through naturally occurring stories told during dinners in white middle class households in Southern California, both mothers and fathers participated in replicating male dominance (the "father knows best" syndrome) by the distribution of participant roles such as protagonist (often a child but sometimes mother and almost never the father) and "problematizer" (often the father, who raised uncomfortable questions or challenged the competence of the protagonist). When mothers collaborated

with children to get their stories told, they unwittingly set themselves up to be subject to this process.

Schieffelin's more recent research has uncovered the socializing role of pastors and other fairly new Bosavi converts in the Southern Highlands, Papua New Guinea community she studies. Pastors have introduced new ways of conveying knowledge, new linguistic epistemic markers—and new ways of speaking about time. And they have struggled with and largely resisted those parts of the Bible that speak of being able to know the inner states of others (e.g. the gospel of Mark, chapter 2, verses 6–8).

## **Ideologies**

In a third example of the current (third) paradigm, since Roman Jakobson's student Michael Silverstein opened the way, there has been an increase in the work done by linguistic anthropologists on the major anthropological theme of ideologies,—in this case "language ideologies", sometimes defined as "shared bodies of commonsense notions about the nature of language in the world." Silverstein has demonstrated that these ideologies are not mere false consciousness but actually influence the evolution of linguistic structures, including the dropping of "thee" and "thou" from everyday English usage. Woolard, in her overview of "code switching", or the systematic practice of alternating linguistic varieties within a conversation or even a single utterance, finds the underlying question anthropologists ask of the practice—Why do they do that?—reflects a dominant linguistic ideology. It is the ideology that people should "really" be monoglot and efficiently targeted

toward referential clarity rather than diverting themselves with the messiness of multiple varieties in play at a single time.

Much research on linguistic ideologies probes subtler influences on language, such as the pull exerted on Tewa, a Kiowa-Tanoan language spoken in certain New Mexican pueblos and on the Hopi Reservation in Arizona, by "kiva speech", discussed in the next section.

Other linguists have carried out research in the areas of language contact, language endangerment, and 'English as a global language'. For instance, Indian linguist BrajKachru investigated local varieties of English in South Asia, the ways in which English functions as a lingua franca among multicultural groups in India. British linguist David Crystal has contributed to investigations of language death attention to the effects of cultural assimilation resulting in the spread of one dominant language in situations of colonialism.

## **Heritage language ideologies**

More recently, a new line of ideology work is beginning to enter the field of linguistics in relation to heritage languages. Specifically, applied linguist Martin Guardado has posited that heritage language ideologies are "somewhat fluid sets of understandings, justifications, beliefs, and judgments that linguistic minorities hold about their languages." Guardado goes on to argue that ideologies of heritage languages also contain the expectations and desires of linguistic minority families "regarding the relevance of these languages in their children's lives as well as when, where, how, and to what ends these languages should be used." Although this is arguably a



fledgling line of language ideology research, this work is poised to contribute to the understanding of how ideologies of language operate in a variety of settings.

## **Social space**

In a final example of this third paradigm, a group of linguistic anthropologists have done very creative work on the idea of social space. Duranti published a groundbreaking article on Samoan greetings and their use and transformation of social space. Before that, Indonesianist Joseph Errington, making use of earlier work by Indonesianists not necessarily concerned with language issues per se, brought linguistic anthropological methods (and semiotic theory) to bear on the notion of the exemplary center, the center of political and ritual power from which emanated exemplary behavior. Errington demonstrated how the Javanese \*priyayi\*, whose ancestors served at the Javanese royal courts, became emissaries, so to speak, long after those courts had ceased to exist, representing throughout Java the highest example of "refined speech." The work of Joel Kuipers develops this theme vis-a-vis the island of Sumba, Indonesia. And, even though it pertains to Tewa Indians in Arizona rather than Indonesians, Paul Kroskrity's argument that speech forms originating in the Tewa kiva (or underground ceremonial space) forms the dominant model for all Tewa speech can be seen as a direct parallel.

Silverstein tries to find the maximum theoretical significance and applicability in this idea of exemplary centers. He feels, in fact, that the exemplary center idea is one of linguistic anthropology's three most important findings. He generalizes the notion thus, arguing "there are wider-scale institutional

'orders of interactionality,' historically contingent yet structured. Within such large-scale, macrosocial orders, in-effect ritual centers of semiosis come to exert a structuring, value-conferring influence on any particular event of discursive interaction with respect to the meanings and significance of the verbal and other semiotic forms used in it." Current approaches to such classic anthropological topics as ritual by linguistic anthropologists emphasize not static linguistic structures but the unfolding in realtime of a "'hypertrophic' set of parallel orders of iconicity and indexicality that seem to cause the ritual to create its own sacred space through what appears, often, to be the magic of textual and nontextualmetricalizations, synchronized."

## **Race, class, and gender**

Addressing the broad central concerns of the subfield and drawing from its core theories, many scholars focus on the intersections of language and the particularly salient social constructs of race (and ethnicity), class, and gender (and sexuality). These works generally consider the roles of social structures (e.g., ideologies and institutions) related to race, class, and gender (e.g., marriage, labor, pop culture, education) in terms of their constructions and in terms of individuals' lived experiences. A short list of linguistic anthropological texts that address these topics follows:

## **Race and ethnicity**

- Alim, H. Samy, John R. Rickford, and Arnetha F. Ball. 2016. *Raciolinguistics: How Language Shapes Our Ideas about Race*. Oxford University Press.
- Bucholtz, Mary. 2001. "The Whiteness of Nerds: Superstandard English and Racial Markedness." *Journal of Linguistic Anthropology* 11 (1): 84–100. doi:10.1525/jlin.2001.11.1.84.
- Bucholtz, Mary. 2010. *White Kids: Language, Race, and Styles of Youth Identity*. Cambridge University Press.
- Davis, Jenny L. 2018. *Talking Indian: Identity and Language Revitalization in the Chickasaw Renaissance*. University of Arizona Press.
- Dick, H. 2011. "Making Immigrants Illegal in Small-Town USA." *Journal of Linguistic Anthropology*. 21(S1):E35-E55.
- Hill, Jane H. 1998. "Language, Race, and White Public Space." *American Anthropologist* 100 (3): 680–89. doi:10.1525/aa.1998.100.3.680.
- Hill, Jane H. 2008. *The Everyday Language of White Racism*. Wiley-Blackwell.
- García-Sánchez, Inmaculada M. 2014. *Language and Muslim Immigrant Childhoods: The Politics of Belonging*. John Wiley & Sons.
- Ibrahim, Awad. 2014. *The Rhizome of Blackness: A Critical Ethnography of Hip-Hop Culture, Language, Identity, and the Politics of Becoming*. 1 edition. New York: Peter Lang Publishing Inc.

- Rosa, Jonathan. 2019. *Looking like a Language, Sounding like a Race: Raciolinguistic Ideologies and the Learning of Latinidad*. Oxford University Press.
- Smalls, Krystal. 2018. "Fighting Words: Antiblackness and Discursive Violence in an American High School." *Journal of Linguistic Anthropology*. 23(3):356-383.
- Spears, Arthur Kean. 1999. *Race and Ideology: Language, Symbolism, and Popular Culture*. Wayne State University Press.
- Urciuoli, Bonnie. 2013. *Exposing Prejudice: Puerto Rican Experiences of Language, Race, and Class*. Waveland Press.
- Wirtz, Kristina. 2011. "Cuban Performances of Blackness as the Timeless Past Still Among Us." *Journal of Linguistic Anthropology*. 21(S1):E11-E34.

## **Class**

- Fox, Aaron A. 2004. *Real Country: Music and Language in Working-Class Culture*. Duke University Press.
- Shankar, Shalini. 2008. *Desi Land: Teen Culture, Class, and Success in Silicon Valley*. Duke University Press.
- Nakassis, Constantine V. 2016. *Doing Style: Youth and Mass Mediation in South India*. University of Chicago Press.

## **Gender and sexuality**

- Bucholtz, Mary. 1999. "'Why be normal?': Language and Identity Practices in a Community of Nerd Girls". *Language in Society*. 28 (2): 207–210.
- Fader, Ayala. 2009. *Mitzvah Girls: Bringing Up the Next Generation of Hasidic Jews in Brooklyn*. Princeton University Press.
- Gaudio, Rudolf Pell. 2011. *Allah Made Us: Sexual Outlaws in an Islamic African City*. John Wiley & Sons.
- Hall, Kira, and Mary Bucholtz. 1995. *Gender Articulated: Language and the Socially Constructed Self*. New York: Routledge.
- Jacobs-Huey, Lanita. 2006. *From the Kitchen to the Parlor: Language and Becoming in African American Women's Hair Care*. Oxford University Press.
- Kulick, Don. 2000. "Gay and Lesbian Language." *Annual Review of Anthropology* 29 (1): 243–85. doi:10.1146/annurev.anthro.29.1.243.
- Kulick, Don. 2008. "Gender Politics." *Men and Masculinities* 11 (2): 186–92. doi:10.1177/1097184X08315098.
- Kulick, Don. 1997. "The Gender of Brazilian Transgendered Prostitutes." *American Anthropologist* 99 (3): 574–85.
- Livia, Anna, and Kira Hall. 1997. *Queerly Phrased: Language, Gender, and Sexuality*. Oxford University Press.
- Manalansan, Martin F. IV. "'Performing' the Filipino Gay Experiences in America: Linguistic Strategies in

a Transnational Context." *Beyond the Lavender Lexicon: Authenticity, Imagination and Appropriation in Lesbian and Gay Language*. Ed. William L Leap. New York: Gordon and Breach, 1997. 249–266

- Mendoza-Denton, Norma. 2014. *Homegirls: Language and Cultural Practice Among Latina Youth Gangs*. John Wiley & Sons.
- Rampton, Ben. 1995. *Crossing: Language and Ethnicity Among Adolescents*. Longman.
- Zimman, Lal, Jenny L. Davis, and Joshua Raclaw. 2014. *Queer Excursions: Rethorizing Binaries in Language, Gender, and Sexuality*. Oxford University Press.

## **Ethnopoetics**

Ethnopoetics is a method of recording text versions of oral poetry or narrative performances (i.e. verbal lore) that uses poetic lines, verses, and stanzas (instead of prose paragraphs) to capture the formal, poetic performance elements which would otherwise be lost in the written texts. The goal of any ethnopoetic text is to show how the techniques of unique oral performers enhance the aesthetic value of their performances within their specific cultural contexts. Major contributors to ethnopoetic theory include Jerome Rothenberg, Dennis Tedlock, and Dell Hymes. Ethnopoetics is considered a subfield of ethnology, anthropology, folkloristics, stylistics, linguistics, and literature and translation studies.

## **Endangered languages: Language documentation and revitalization**

Endangered languages are languages that are not being passed down to children as their mother tongue or that have declining numbers of speakers for a variety of reasons. Therefore, after a couple generations these languages may no longer be spoken. Anthropologists have been involved with endangered language communities through their involvement in language documentation and revitalization projects.

In a language documentation project, researchers work to develop records of the language - these records could be field notes and audio or video recordings. To follow best practices of documentation, these records should be clearly annotated and kept safe within an archive of some kind. Franz Boas was one of the first anthropologists involved in language documentation within North America and he supported the development of three key materials: 1) grammars, 2) texts, and 3) dictionaries. This is now known as the Boasian Trilogy.

Language revitalization is the practice of bringing a language back into common use. The revitalization efforts can take the form of teaching the language to new speakers or encouraging the continued use within the community. One example of a language revitalization project is the Lenape language course taught at Swathmore College, Pennsylvania. The course aims to educate indigenous and non-indigenous students about the Lenape language and culture.

Language reclamation, as a subset of revitalization, implies that a language has been taken away from a community and addresses their concern in taking back the agency to revitalize their language on their own terms. Language reclamation addresses the power dynamics associated with language loss. Encouraging those who already know the language to use it, increasing the domains of usage, and increasing the overall prestige of the language are all components of reclamation. One example of this is the Miami language being brought back from 'extinct' status through extensive archives.

While the field of linguistics has also been focused on the study of the linguistic structures of endangered languages, anthropologists also contribute to this field through their emphasis on ethnographic understandings of the socio-historical context of language endangerment, but also of language revitalization and reclamation projects.



## Chapter 3

# Language Revitalization

**Language revitalization**, also referred to as **language revival** or **reversing language shift**, is an attempt to halt or reverse the decline of a language or to revive an extinct one. Those involved can include parties such as linguists, cultural or community groups, or governments. Some argue for a distinction between **language revival** (the resurrection of an extinct language with no existing native speakers) and **language revitalization** (the rescue of a "dying" language). It has been pointed out that there has only been one successful instance of a complete language revival, that of the Hebrew language, creating a new generation of native speakers without any pre-existing native speakers as a model.

Languages targeted for language revitalization include those whose use and prominence is severely limited. Sometimes various tactics of language revitalization can even be used to try to revive extinct languages. Though the goals of language revitalization vary greatly from case to case, they typically involve attempting to expand the number of speakers and use of a language, or trying to maintain the current level of use to protect the language from extinction or language death.

Reasons for revitalization vary. In recent times alone, it is estimated that more than 2000 languages have already become extinct. The UN estimates that more than half of the languages spoken today have fewer than 10,000 speakers and that a quarter have fewer than 1,000 speakers and that, unless there

are some efforts to maintain them, over the next hundred years most of these will become extinct. These figures are often cited as reasons why language revitalization is necessary to preserve linguistic diversity. Culture and identity are also frequently cited reasons for language revitalization, when a language is perceived as a unique "cultural treasure." A community often sees language as a unique part of their culture, connecting them with their ancestors or with the land, making up an essential part of their history and self-image.

Language revitalization is also closely tied to the linguistic field of language documentation. In this field, linguists attempt to create full records of a language's grammar, vocabulary, and linguistic features. This practice can often lead to more concern for the revitalization of a specific language on study. Furthermore, the task of documentation is often taken on with the goal of revitalization in mind.

## **The five degrees of language endangerment**

- Healthy/strong
- all generations use language in variety of settings
- Weakening/sick
- spoken by older people; not fully used in younger generations
- Moribund/dying
- only a few speakers (non-children) remain; no longer used as native language by children
- Dead

- no longer spoken as a native language
- Extinct
- no longer spoken and barely has written records (if at all)

Another scale for identifying degrees of language endangerment is used in a 2003 paper ("Language Vitality and Endangerment") commissioned by UNESCO from an international group of linguists. The linguists, among other goals and priorities, create a scale with six degrees for language vitality and endangerment. They also propose nine factors (six of which use the six-degree scale) to "characterizing a language's overall sociolinguistic situation". The nine factors with their respective scales are:

- Intergenerational language transmission
- safe: all generations use the language
- unsafe: some children use the language in all settings, all children use the language in some settings
- definitively endangered: few children speak the language; predominantly spoken by the parental generation and older
- severely endangered: spoken by older generations; not used by the parental generation and younger
- critically endangered: few speakers remain and are mainly from the great grandparental generation
- extinct: no living speakers
- Absolute number of speakers
- Proportion of speakers within the total population
- safe: the language is spoken by 100% of the population

- unsafe: the language is spoken by nearly 100% of the population
- definitively endangered: the language is spoken by a majority of the population
- severely endangered: the language is spoken by less than 50% of the population
- critically endangered: the language has very few speakers
- extinct: no living speakers
- Trends in existing language domains
- universal use (safe): spoken in all domains; for all functions
- multilingual parity (unsafe): multiple languages (2+) are spoken in most social domains; for most functions
- dwindling domains (definitively endangered): mainly spoken in home domains and is in competition with the dominant language; for many functions
- limited or formal domains (severely endangered): spoken in limited social domains; for several functions
- highly limited domains (critically endangered): spoken in highly restricted domains; for minimal functions
- extinct: no domains; no functions
- Response to new domains and media
- dynamic (safe): spoken in all new domains
- robust/active (unsafe): spoken in most new domains
- receptive (definitively endangered): spoken in many new domains
- coping (severely endangered): spoken in some new domains

- minimal (critically endangered): spoken in minimal new domains
- inactive (extinct): spoken in no new domains
- Materials for language education and literacy
- safe: established orthography and extensive access to educational materials
- unsafe: access to educational materials; children developing literacy; not used by administration
- definitively endangered: access to educational materials exist at school; literacy in language is not promoted
- severely endangered: literacy materials exist however are not present in school curriculum
- critically endangered: orthography is known and some written materials exist
- extinct: no orthography is known
- Governmental and institutional language attitudes and policies (including official status and use)
- equal support (safe): all languages are equally protected
- differentiated support (unsafe): primarily protected for private domains
- passive assimilation (definitively endangered): no explicit protective policy; language use dwindles in public domain
- active assimilation (severely endangered): government discourages use of language; no governmental protection of language in any domain
- forced assimilation (critically endangered): language is not recognized or protected; government recognized another official language
- prohibition (extinct): use of language is banned

- Community members' attitudes towards their own language
- safe: language is revered, valued, and promoted by whole community
- unsafe: language maintenance is supported by most of the community
- definitively endangered: language maintenance is supported by much of the community; the rest are indifferent or support language loss
- severely endangered: language maintenance is supported by some of the community; the rest are indifferent or support language loss
- critically endangered: language maintenance is supported by only a few members of the community; the rest are indifferent or support language loss
- extinct: complete apathy towards language maintenance; prefer dominant language
- Amount and quality of documentation.
- superlative (safe): extensive audio, video, media, and written documentation of the language
- good (unsafe): audio, video, media, and written documentation all exist; a handful of each
- fair (definitively endangered): some audio and video documentation exists; adequate written documentation
- fragmentary (severely endangered): minimal audio and video documentation exists at low quality; minimal written documentation
- inadequate (critically endangered): only a handful of written documentation exists
- undocumented (extinct): no documentation exists

# **Theory**

One of the most important preliminary steps in language revitalization/recovering involves establishing the degree to which a particular language has been “dislocated”. This helps involved parties find the best way to assist or revive the language.

## **Steps in reversing language shift**

There are many different theories or models that attempt to lay out a plan for language revitalization. One of these is provided by celebrated linguist Joshua Fishman. Fishman's model for reviving threatened (or sleeping) languages, or for making them sustainable, consists of an eight-stage process. Efforts should be concentrated on the earlier stages of restoration until they have been consolidated before proceeding to the later stages. The eight stages are:

- Acquisition of the language by adults, who in effect act as language apprentices (recommended where most of the remaining speakers of the language are elderly and socially isolated from other speakers of the language).
- Create a socially integrated population of active speakers (or users) of the language (at this stage it is usually best to concentrate mainly on the spoken language rather than the written language).
- In localities where there are a reasonable number of people habitually using the language, encourage the informal use of the language among people of all age

groups and within families and bolster its daily use through the establishment of local neighbourhood institutions in which the language is encouraged, protected and (in certain contexts at least) used exclusively.

- In areas where oral competence in the language has been achieved in all age groups, encourage literacy in the language, but in a way that does not depend upon assistance from (or goodwill of) the state education system.
- Where the state permits it, and where numbers warrant, encourage the use of the language in compulsory state education.
- Where the above stages have been achieved and consolidated, encourage the use of the language in the workplace.
- Where the above stages have been achieved and consolidated, encourage the use of the language in local government services and mass media.
- Where the above stages have been achieved and consolidated, encourage use of the language in higher education, government, etc.

This model of language revival is intended to direct efforts to where they are most effective and to avoid wasting energy trying to achieve the later stages of recovery when the earlier stages have not been achieved. For instance, it is probably wasteful to campaign for the use of a language on television or in government services if hardly any families are in the habit of using the language.



Additionally, TasakuTsunoda describes a range of different techniques or methods that speakers can use to try to revitalize a language, including techniques to revive extinct languages and maintain weak ones. The techniques he lists are often limited to the current vitality of the language.

He claims that the immersion method cannot be used to revitalize an extinct or moribund language. In contrast, the master-apprentice method of one-on-one transmission on language proficiency can be used with moribund languages. Several other methods of revitalization, including those that rely on technology such as recordings or media, can be used for languages in any state of viability.

## **Factors in successful language revitalization**

David Crystal, in his book *Language Death*, proposes that language revitalization is more likely to be successful if its speakers

- increase the language's prestige within the dominant community;
- increase their wealth and income;
- increase their legitimate power in the eyes of the dominant community;
- have a strong presence in the education system;
- can write down the language;
- can use electronic technology.

In her book, *Endangered Languages: An Introduction*, Sarah Thomason notes the success of revival efforts for modern Hebrew and the relative success of revitalizing Maori in New Zealand (see *Specific Examples* below). One notable factor these two examples share is that the children were raised in fully immersive environments. In the case of Hebrew, it was on early collective-communities called kibbutzim. For the Maori language In New Zealand, this was done through a language nest.

## **Revival linguistics**

Ghil'ad Zuckermann proposes "Revival Linguistics" as a new linguistic discipline and paradigm.

Zuckermann's term 'Revival Linguistics' is modelled upon 'Contact Linguistics'. Revival linguistics inter alia explores the universal constraints and mechanisms involved in language reclamation, renewal and revitalization. It draws perspicacious comparative insights from one revival attempt to another, thus acting as an epistemological bridge between parallel discourses in various local attempts to revive sleeping tongues all over the globe.

According to Zuckermann, "revival linguistics combines scientific studies of native language acquisition and foreign language learning. After all, language reclamation is the most extreme case of second-language learning. Revival linguistics complements the established area of documentary linguistics, which records endangered languages before they fall asleep."

Zuckermann proposes that "revival linguistics changes the field of historical linguistics by, for instance, weakening the family tree model, which implies that a language has only one parent."

There are disagreements in the field of language revitalization as to the degree that revival should concentrate on maintaining the traditional language, versus allowing simplification or widespread borrowing from the majority language.

## **Compromise**

Zuckermann acknowledges the presence of "local peculiarities and idiosyncrasies" but suggests that "there are linguistic constraints applicable to all revival attempts. Mastering them would help revivalists and first nations' leaders to work more efficiently. For example, it is easier to resurrect basic vocabulary and verbal conjugations than sounds and word order. Revivalists should be realistic and abandon discouraging, counter-productive slogans such as "Give us authenticity or give us death!"

Nancy Dorian has pointed out that conservative attitudes toward loanwords and grammatical changes often hamper efforts to revitalize endangered languages (as with Tiwi in Australia), and that a division can exist between educated revitalizers, interested in historicity, and remaining speakers interested in locally authentic idiom (as has sometimes occurred with Irish). Some have argued that structural compromise may, in fact, enhance the prospects of survival, as

may have been the case with English in the post-Norman period.

## **Traditionalist**

Other linguists have argued that when language revitalization borrows heavily from the majority language, the result is a new language, perhaps a creole or pidgin. For example, the existence of "Neo-Hawaiian" as a separate language from "Traditional Hawaiian" has been proposed, due to the heavy influence of English on every aspect of the revived Hawaiian language. This has also been proposed for Irish, with a sharp division between "Urban Irish" (spoken by second-language speakers) and traditional Irish (as spoken as a first language in Gaeltacht areas). Ó Béarrastated: "[to] follow the syntax and idiomatic conventions of English, [would be] producing what amounts to little more than English in Irish drag." With regard to the then-moribund Manx language, the scholar T. F. O'Rahilly stated, "When a language surrenders itself to foreign idiom, and when all its speakers become bilingual, the penalty is death." Neil McRae has stated that the uses of Scottish Gaelic are becoming increasingly tokenistic, and native Gaelic idiom is being lost in favor of artificial terms created by second-language speakers.

## **Specific examples**

The total revival of a dead language (in the sense of having no native speakers) to become the shared means of communication of a self-sustaining community of several

million first language speakers has happened only once, in the case of Hebrew, now the national language of Israel. In this case, there was a unique set of historical and cultural characteristics that facilitated the revival. (See Revival of the Hebrew language.) Hebrew, once largely a liturgical language, was re-established as a means of everyday communication by Jews migrating to what is now the State of Israel and the Palestinian territories, starting in the nineteenth century. It is the world's most famous and successful example of language revitalization.

In a related development, literary languages without native speakers enjoyed great prestige and practical utility as *lingua francas*, often counting millions of fluent speakers at a time. In many such cases, a decline in the use of the literary language, sometimes precipitous, was later accompanied by a strong renewal. This happened, for example, in the revival of Classical Latin in the Renaissance, and the revival of Sanskrit in the early centuries AD. An analogous phenomenon in contemporary Arabic-speaking areas is the expanded use of the literary language (Modern Standard Arabic, a form of the Classical Arabic of the 6th century AD). This is taught to all educated speakers and is used in radio broadcasts, formal discussions, etc.

In addition, literary languages have sometimes risen to the level of becoming first languages of very large language communities. An example is standard Italian, which originated as a literary language based on the language of 13th-century Florence, especially as used by such important Florentine writers as Dante, Petrarch and Boccaccio. This language existed for several centuries primarily as a literary vehicle,

with few native speakers; even as late as 1861, on the eve of Italian unification, the language only counted about 500,000 speakers (many non-native), out of a total population of c. 22,000,000. The subsequent success of the language has been through conscious development, where speakers of any of the numerous Italian languages were taught standard Italian as a second language and subsequently imparted it to their children, who learned it as a first language. Of course this came at the expense of local Italian languages, most of which are now endangered. Success was enjoyed in similar circumstances by High German, standard Czech, Castilian Spanish and other languages.

## **Africa**

The Coptic language began its decline when Arabic became the predominant language in Egypt. Pope Shenouda III established the Coptic Language Institute in December 1976 in Saint Mark's Coptic Orthodox Cathedral in Cairo for the purpose of reviving the Coptic language.

## **Americas**

### **North America**

In recent years, a growing number of Native American tribes have been trying to revitalize their languages. For example, there is an Apple iPhone/iPod app for the Halq'emeylem language of the Greater Vancouver region of Canada. In addition, there are apps (including phrases, word lists and

dictionaries) in many Native languages ranging from Cree, Cherokee and Chickasaw, to Lakota, Ojibway and Oneida, Massachusett, Navajo and Gwyh'in.

Wampanoag, a language spoken by the people of the same name in Massachusetts, underwent a language revival project led by Jessie Little Doe Baird, a trained linguist. Members of the tribe use the extensive written records that exist in their language, including a translation of the Bible and legal documents, in order to learn and teach Wampanoag. The project has seen children speaking the language fluently for the first time in over 100 years. In addition, there are currently attempts at reviving the Chochenyo language of California, which had become extinct.

## **Tlingit**

Similar to other Indigenous languages, Tlingit is critically endangered. Less than 100 fluent Elders continue to exist. From 2013 to 2014, the language activist, author, and teacher, Sʔimlaʔxw Michele K. Johnson from the Syilx Nation, attempted to teach two hopeful learners of Tlingit in the Yukon. Her methods included textbook creation, sequenced immersion curriculum, and film assessment. The aim was to assist in the creation of adult speakers that are of parent-age, so that they too can begin teaching the language. In 2020, X̱'unei Lance Twitchell led an Tlingit online class with Outer Coast College. Dozens of students participated. He is an associate professor of Alaska Native Languages in the School of Arts and Sciences at the University of Alaska Southeast which offers a minor in Tlingit language and an emphasis on Alaska

Native Languages and Studies within a Bachelor's degree in Liberal Arts.

## **South America**

Kichwa is the variety of the Quechua language spoken in Ecuador and is one of the most widely spoken indigenous languages in South America. Despite this fact, Kichwa is a threatened language, mainly because of the expansion of Spanish in South America. One community of original Kichwa speakers, Lagunas, was one of the first indigenous communities to switch to the Spanish language. According to King, this was because of the increase of trade and business with the large Spanish-speaking town nearby. The Lagunas people assert that it was not for cultural assimilation purposes, as they value their cultural identity highly. However, once this contact was made, language for the Lagunas people shifted through generations, to Kichwa and Spanish bilingualism and now is essentially Spanish monolingualism. The feelings of the Lagunas people present a dichotomy with language use, as most of the Lagunas members speak Spanish exclusively and only know a few words in Kichwa.

The prospects for Kichwa language revitalization are not promising, as parents depend on schooling for this purpose, which is not nearly as effective as continual language exposure in the home. Schooling in the Lagunas community, although having a conscious focus on teaching Kichwa, consists of mainly passive interaction, reading, and writing in Kichwa. In addition to grassroots efforts, national language revitalization organizations, like CONAIE, focus attention on non-Spanish



speaking indigenous children, who represent a large minority in the country. Another national initiative, Bilingual Intercultural Education Project (PEBI), was ineffective in language revitalization because instruction was given in Kichwa and Spanish was taught as a second language to children who were almost exclusively Spanish monolinguals. Although some techniques seem ineffective, Kendall A. King provides several suggestions:

- Exposure to and acquisition of the language at a young age.
- Extreme immersion techniques.
- Multiple and diverse efforts to reach adults.
- Flexibility and coordination in planning and implementation
- Directly addressing different varieties of the language.
- Planners stressing that language revitalization is a long process
- Involving as many people as possible
- Parents using the language with their children
- Planners and advocates approaching the problem from all directions.

Specific suggestions include imparting an elevated perception of the language in schools, focusing on grassroots efforts both in school and the home, and maintaining national and regional attention.

## **Asia**

### **Hebrew**

The revival of the Hebrew language is the only truly successful example of a revived dead language. The Hebrew language survived into the medieval period as the language of Jewish liturgy and rabbinic literature. With the rise of Zionism in the 19th century, it was revived as a spoken and literary language, becoming primarily a spoken lingua franca among the early Jewish immigrants to Ottoman Palestine and received the official status in the 1922 constitution of the British Mandate for Palestine and subsequently of the State of Israel.

### **Sanskrit**

The revival of Sanskrit happened in India. In the 2001 census of India, 14,135 people claimed Sanskrit as their mother tongue. It increased to 24,821 people in the 2011 census of India. Sanskrit has experienced a recorded a growth of over 70 per cent in one decade due to the Sanskrit revival. Many Sanskrit speaking villages were also developed. However, Sanskrit speakers still account for just 0.00198 percent of India's total population.

### **Soyot**

The Soyot language of the small-numbered Soyots in Buryatia, Russia, one of Siberian Turkic languages, has been

reconstructed and a Soyot-Buryat-Russian dictionary has been published in 2002. The language is currently taught in some elementary schools.

## **Ainu**

The Ainu language of the indigenous Ainu people of northern Japan is currently moribund, but efforts are underway to revive it. A 2006 survey of the Hokkaido Ainu indicated that only 4.6% of Ainu surveyed were able to converse in or "speak a little" Ainu. As of 2001, Ainu was not taught in any elementary or secondary schools in Japan, but was offered at numerous language centres and universities in Hokkaido, as well as at Tokyo's Chiba University.

## **Manchu**

In China, the Manchu language is one of the most endangered languages, with speakers only in three small areas of Manchuria remaining. Some enthusiasts are trying to revive the language of their ancestors using available dictionaries and textbooks, and even occasional visits to QapqalXibe Autonomous County in Xinjiang, where the related Xibe language is still spoken natively.

## **Philippine Spanish**

In the Philippines, a variation of Spanish that was primarily based on Mexican Spanish was the *lingua franca* of the country

since Spanish colonization in 1565 and was an official language alongside Filipino (standardized Tagalog) and English until 1987, following a ratification of a new constitution, where it was re-designated as a voluntary language. As a result of its loss as an official language and years of marginalization at the official level during and after American colonization, the use of Spanish amongst the overall populace decreased dramatically and became moribund, with the remaining native speakers left being mostly elderly people. However, it is currently seeing a slow revival due to past government promotion under the administration of former President Gloria Macapagal Arroyo. Most notably, Resolution No. 2006-028 reinstated Spanish as a mandatory subject in secondary schools and universities. Results were immediate as the job demand for Spanish speakers had increased since 2008. As of 2010, the Instituto Cervantes in Manila reported the number of Filipino Hispanophones with native or non-native knowledge at approximately 3 million (including those who speak the Spanish-based creole Chavacano). In addition to government efforts, Spanish has also seen a small revival of interest in media, thanks to the importing of telenovelas and music from Latin America.

## **Other Asian**

The KodrahKristang revitalization initiative in Singapore seeks to revive the critically endangered Kristang creole.

In Thailand, there exists a Chong language revitalization project, headed by SuwilaiPremsrirat.

## **Europe**

In Europe, in the 19th and early 20th centuries, the use of both local and learned languages declined as the central governments of the different states imposed their vernacular language as the standard throughout education and official use (this was the case in the United Kingdom, France, Spain, Italy and Greece, and to some extent, in Germany and Austria-Hungary).

In the last few decades, local nationalism and human rights movements have made a more multicultural policy standard in European states; sharp condemnation of the earlier practices of suppressing regional languages was expressed in the use of such terms as "linguicide".

## **Irish**

One of the best known European attempts at language revitalization concerns the Irish language. While English is dominant through most of Ireland, Irish, a Celtic language, is still spoken in certain areas called *Gaeltachtaí*, but there it is in serious decline. The challenges faced by the language over the last few centuries have included exclusion from important domains, social denigration, the death or emigration of many Irish speakers during the Irish famine of the 1840s, and continued emigration since. Efforts to revitalise Irish were being made, however, from the mid-1800s, and were associated with a desire for Irish political independence. Contemporary Irish language revitalization has chiefly involved teaching Irish as a compulsory language in mainstream English-speaking

schools. But the failure to teach it in an effective and engaging way means (as linguist Andrew Carnie notes) that students do not acquire the fluency needed for the lasting viability of the language, and this leads to boredom and resentment. Carnie also noted a lack of media in Irish (2006), though this is no longer the case.

The decline of the *Gaeltachtaí* and the failure of state-directed revitalization have been countered by an urban revival movement. This is largely based on an independent community-based school system, known generally as *Gaelscoileanna*. These schools teach entirely through Irish and their number is growing, with over thirty such schools in Dublin alone. They are an important element in the creation of a network of urban Irish speakers (known as *Gaeilgeoirí*), who tend to be young, well-educated and middle-class. It is now likely that this group has acquired critical mass, a fact reflected in the expansion of Irish-language media. Irish language television has enjoyed particular success. It has been argued that they tend to be better educated than monolingual English speakers and enjoy higher social status. They represent the transition of Irish to a modern urban world, with an accompanying rise in prestige.

## **Scottish Gaelic**

There are also current attempts to revive the related language of Scottish Gaelic, which was suppressed following the formation of the United Kingdom, and entered further decline due to the Highland clearances. Currently, Gaelic is only spoken widely in the Western Isles and some relatively small

areas of the Highlands and Islands. The decline in fluent Gaelic speakers has slowed; however, the population center has shifted to L2 speakers in urban areas, especially Glasgow.

## **Manx**

Another Celtic language, Manx, lost its last native speaker in 1974 and was declared extinct by UNESCO in 2009, but never completely fell from use. The language is now taught in primary and secondary schools, including as a teaching medium at the BunscoillGhaelgagh, used in some public events and spoken as a second language by approximately 1800 people. Revitalization efforts include radio shows in Manx Gaelic and social media and online resources. The Manx government has also been involved in the effort by creating organizations such as the Manx Heritage Foundation (Culture Vannin) and the position of Manx Language Officer. The government has released an official Manx Language Strategy for 2017–2021.

## **Cornish**

There have been a number of attempts to revive the Cornish language, both privately and some under the Cornish Language Partnership. Some of the activities have included translation of the Christian scriptures, a guild of bards, and the promotion of Cornish literature in modern Cornish, including novels and poetry.

## **Caló**

The Romani arriving in the Iberian Peninsula developed an Iberian Romani dialect. As time passed, Romani ceased to be a full language and became Caló, a cant mixing Iberian Romance grammar and Romani vocabulary. With sedentarization and obligatory instruction in the official languages, Calóis used less and less. As Iberian Romani proper is extinct and as Calóis endangered, some people are trying to revitalise the language. The Spanish politician Juan de Dios Ramírez Heredia promotes Romanò-Kalò, a variant of International Romani, enriched by Caló words. His goal is to reunify the Caló and Romani roots.

## **Livonian**

The Livonian language, a Finnic language, once spoken on about a third of modern-day Latvian territory, died in the 21st century with the death of the last native speaker Grizelda Kristiņa on 2 June 2013. Today there are about 210 people mainly living in Latvia who identify themselves as Livonian and speak the language on the A1-A2 level according to the Common European Framework of Reference for Languages and between 20 and 40 people who speak the language on level B1 and up. Today all speakers learn Livonian as a second language. There are different programs educating Latvians on the cultural and linguistic heritage of Livonians and the fact that most Latvians have common Livonian descent.



Programs worth mentioning include:

- Livones.net with extensive information about language, history and culture
- The Livonian Institute of the University of Latvia doing research on the Livonian language, other Finnic languages in Latvia and providing an extensive Livonian-Latvian-Estonian dictionary with declinations/conjugations
- Virtual Livonia providing information on the Livonian language and especially its grammar
- Mierlinkizt: An annual summer camp for children to teach children about the Livonian language, culture etc.
- Līvōdīt (Livonian Union)

The Livonian linguistic and cultural heritage is included in the Latvian cultural canon and the protection, revitalization and development of Livonian as an indigenous language is guaranteed by Latvian law

## **Old Prussian**

A few linguists and philologists are involved in reviving a reconstructed form of the extinct Old Prussian language from Luther's catechisms, the Elbing Vocabulary, place names, and Prussian loanwords in the Low Prussian dialect of German. Several dozen people use the language in Lithuania, Kaliningrad, and Poland, including a few children who are natively bilingual.

The Prusaspirā Society has published their translation of Antoine de Saint-Exupéry's *The Little Prince*. The book was translated by Piotr Szatkowski (PīterisŠātkis) and released in 2015. The other efforts of Baltic Prussian societies include the development of online dictionaries, learning apps and games. There also have been several attempts to produce music with lyrics written in the revived Baltic Prussian language, most notably in the Kaliningrad Oblast by RomoweRikoito, Kellan and ĀustrasLaiwan, but also in Lithuania by Kūlgrinda in their 2005 album *Prūsų Giesmės* (Prussian Hymns), and in Latvia by Rasa Ensemble in 1988 and Valdis Muktupāvels in his 2005 oratorio "Pārcēlātājs Pontifex" featuring several parts sung in Prussian.

Important in this revival was VytautasMažiulis, who died on 11 April 2009, and his pupil LetasPalmaitis, leader of the experiment and author of the website *Prussian Reconstructions*. Two late contributors were PrāncisArelis (PranciškusErelis), Lithuania, and DailūnsRussinis (DailonisRusiņš), Latvia. After them, TwankstasGlabbis from Kaliningrad oblast and NērtiksPamedīns from East-Prussia, now Polish Warmia-Mazuria actively joined.

## **Oceania**

## **Australia**

The European colonization of Australia, and the consequent damage sustained by Aboriginal communities, had a catastrophic effect on indigenous languages especially in the

southeast and south of the country, leaving some with no living traditional native speakers. A number of Aboriginal communities in Victoria and elsewhere are now trying to revive these languages. The work is typically directed by a group of elders and other knowledgeable people, with community language workers doing most of the research and teaching. They analyze the data, develop spelling systems and vocabulary and prepare resources. Decisions are made in collaboration. Some communities employ linguists, and there are also linguists who have worked independently, such as Luise Hercus and Peter K. Austin.

The Victorian Department of Education and Training reported 1,867 student enrollments in 14 schools offering an Aboriginal Languages Program in the state of Victoria in 2018.

The Pertame Project is an example in Central Australia. Pertame, from the country south of Alice Springs, along the Finke River, is a dialect in the Arrernte group of languages. With only 20 fluent speakers left by 2018, the Pertame Project is seeking to retain and revive the language, headed by Pertame elder Christobel Swan. The Diyari language of the far north of South Australia has an active programme under way, with materials available for teaching in schools and the wider community.

## **New Zealand**

One of the best cases of relative success in language revitalization is the case of Māori, also known as *te reo Māori*. It is the ancestral tongue of the indigenous Māori people of

New Zealand and a vehicle for prose narrative, sung poetry, and genealogical recital. The history of the Māori people is taught in *te reo Māori* in sacred learning houses through oral transmission. Even after *te reo Māori* became a written language, the oral tradition was preserved.

Once European colonization began, many laws were enacted in order to promote the use of English over *te reo Māori* among indigenous people. The Education Ordinance Act of 1847 mandated school instruction in English and established boarding schools to speed up assimilation of Māori youths into European culture. The Native School Act of 1858 forbade *te reo Māori* from being spoken in schools. The colonial masters also promoted the use of English in Māori homes, convincing many parents that their children would not get jobs unless they spoke English.

During the 1970s, a group of young Māori people, the NgāTamatoa, successfully campaigned for Māori to be taught in schools. Also, Kohanga Reo, Māori language preschools, called language nests, were established. The emphasis was on teaching children the language at a young age, a very effective strategy for language learning. The Māori Language Commission was formed in 1987, leading to a number of national reforms aimed at revitalizing *te reo Māori*. They include media programs broadcast in *te reo Māori*, undergraduate college programs taught in *te reo Māori*, and an annual Māori language week. Each *iwi*, or tribe, created a language planning program catering to its specific circumstances. These efforts have resulted in a steady increase in children being taught in *te reo Māori* in schools since 1996.

## **Hawaiian**

On six of the seven inhabited islands of Hawaii, Hawaiian was displaced by English and is no longer used as the daily language of communication. The one exception is Ni'ihau, where Hawaiian has never been displaced, has never been endangered, and is still used almost exclusively. Efforts to revive the language have increased in recent decades. Hawaiian language immersion schools are now open to children whose families want to retain (or introduce) Hawaiian language into the next generation. The local National Public Radio station features a short segment titled "Hawaiian word of the day". Additionally, the Sunday editions of the *Honolulu Star-Bulletin* feature a brief article called *Kauakukalahale*, written entirely in Hawaiian by a student.

## **Current revitalization efforts**

Language revitalization efforts are ongoing around the world. Revitalization teams are utilizing modern technologies to increase contact with indigenous languages and to record traditional knowledge.

In Mexico, the Mixtec people's language heavily revolves around the interaction between climate, nature, and what it means for their livelihood. UNESCO's LINKS (Local and Indigenous Knowledge) program recently underwent a project to create a glossary of Mixtec terms and phrases related to climate. UNESCO believes that the traditional knowledge of the Mixtec people via their deep connection with weather

phenomena can provide insight on ways to address climate change. Their intention in creating the glossary is to "facilitate discussions between experts and the holders of traditional knowledge".

In Canada, the Wapikoni Mobile project travels to indigenous communities and provides lessons in film making. Program leaders travel across Canada with mobile audiovisual production units, and aims to provide indigenous youth with a way to connect with their culture through a film topic of their choosing. The Wapikona project submits its films to events around the world as an attempt to spread knowledge of indigenous culture and language.

Of the youth in Rapa Nui (Easter Island), ten percent learn their mother language. The rest of the community has adopted Spanish in order to communicate with the outside world and support its tourism industry. Through a collaboration between UNESCO and the Chilean Corporación Nacional de Desarrollo Indígena, the Department of Rapa Nui Language and Culture at the Lorenzo Baeza Vega School was created. Since 1990, the department has created primary education texts in the Rapa Nui language. In 2017, the Nid Rapa Nui, a non-governmental organization was also created with the goal of establishing a school that teaches courses entirely in Rapa Nui.

# **Health benefits of language revitalization**

Language revitalization has been linked to increased health outcomes for Indigenous communities involved in reclaiming traditional language. Benefits range from improved mental health for community members, increasing connectedness to culture, identity, and a sense of wholeness. Indigenous languages are a core element in the formation of identity, providing pathways for cultural expression, agency, spiritual and ancestral connection. Connection to culture is considered to play an important role in childhood development, and is a UN convention right. Much has been written about the connection between identity and culture being inextricably intertwined in Indigenous cultures around the world. As colonisation and subsequent linguisticide was carried out through policies such as those that created Australia's Stolen Generation have damaged this connection, language revitalization may also play an important role in countering intergenerational trauma that has been caused.

One study in the Barngarla Community in South Australia has been looking holistically at the positive benefits of language reclamation, healing mental and emotional scars, and building connections to community and country that underpin wellness and wholeness. The study identified the Barngarla peoples connection to their language is a strong component of developing a strong cultural and personal identity; the people are as connected to language as they are to culture, and culture is key to their identity. Another study in New South

Wales on the Warlpiri people echoes language as life, that the survival of the language is tied to the survival of the community. Language revival is closely linked to overcoming feelings of shame and fear, which have led to poor health outcomes in the past where speaking traditional language meant the possibility of being removed from family and community. Language reclamation is a form of empowerment and builds strong connections with community and wholeness.

## **Criticism**

John McWhorter has argued that programs to revive indigenous languages will almost never be very effective because of the practical difficulties involved. He also argues that the death of a language does not necessarily mean the death of a culture. Indigenous expression is still possible even when the original language has disappeared, as with Native American groups and as evidenced by the vitality of black American culture in the United States, among people who speak not Yoruba but English. He argues that language death is, ironically, a sign of hitherto isolated peoples migrating and sharing space: "To maintain distinct languages across generations happens only amidst unusually tenacious self-isolation—such as that of the Amish—or brutal segregation".

Kenan Malik has also argued that it is "irrational" to try to preserve all the world's languages, as language death is natural and in many cases inevitable, even with intervention. He proposes that language death improves communication by



ensuring more people speak the same language. This may benefit the economy and reduce conflict.

The protection of minority languages from extinction is often not a concern for speakers of the dominant language. Oftentimes, there is prejudice and deliberate persecution of minority languages, in order to appropriate the cultural and economic capital of minority groups. At other times governments deem that the cost of revitalization programs and creating linguistically diverse materials is too great to take on.

## Chapter 4

# Internet Linguistics

**Internet linguistics** is a domain of linguistics advocated by the English linguist David Crystal. It studies new language styles and forms that have arisen under the influence of the Internet and of other new media, such as Short Message Service (SMS) text messaging. Since the beginning of human-computer interaction (HCI) leading to computer-mediated communication (CMC) and Internet-mediated communication (IMC), experts, such as Gretchen McCulloch have acknowledged that linguistics has a contributing role in it, in terms of web interface and usability. Studying the emerging language on the Internet can help improve conceptual organization, translation and web usability. Such study aims to benefit both linguists and web users combined.

The study of Internet linguistics can take place through four main perspectives: sociolinguistics, education, stylistics and applied linguistics. Further dimensions have developed as a result of further technological advances – which include the development of the Web as corpus and the spread and influence of the stylistic variations brought forth by the spread of the Internet, through the mass media and through literary works. In view of the increasing number of users connected to the Internet, the linguistics future of the Internet remains to be determined, as new computer-mediated technologies continue to emerge and people adapt their languages to suit these new media. The Internet continues to play a significant

role both in encouraging people as well as in diverting attention away from the usage of languages.

## **Main perspectives**

David Crystal has identified four main perspectives for further investigation – the sociolinguistic perspective, the educational perspective, the stylistic perspective and the applied perspective. The four perspectives are effectively interlinked and affect one another.

### **Sociolinguistic perspective**

This perspective deals with how society views the impact of Internet development on languages. The advent of the Internet has revolutionized communication in many ways; it changed the way people communicate and created new platforms with far-reaching social impact. Significant avenues include but are not limited to SMS text messaging, e-mails, chat groups, virtual worlds and the Web.

The evolution of these new mediums of communications has raised much concern with regards to the way language is being used. According to Crystal (2005), these concerns are neither without grounds nor unseen in history – it surfaces almost always when a new technology breakthrough influences languages; as seen in the 15th century when printing was introduced, the 19th century when the telephone was invented and the 20th century when broadcasting began to penetrate our society.

At a personal level, CMC such as SMS Text Messaging and mobile e-mailing (push mail) has greatly enhanced instantaneous communication. Some examples include the iPhone and the BlackBerry.

In schools, it is not uncommon for educators and students to be given personalized school e-mail accounts for communication and interaction purposes. Classroom discussions are increasingly being brought onto the Internet in the form of discussion forums. For instance, at Nanyang Technological University, students engage in collaborative learning at the university's portal – edveNTUre, where they participate in discussions on forums and online quizzes and view streaming podcasts prepared by their course instructors among others. iTunes U in 2008 began to collaborate with universities as they converted the Apple music service into a store that makes available academic lectures and scholastic materials for free – they have partnered more than 600 institutions in 18 countries including Oxford, Cambridge and Yale Universities.

These forms of academic social networking and media are slated to rise as educators from all over the world continue to seek new ways to better engage students. It is commonplace for students in New York University to interact with “guest speakers weighing in via Skype, library staffs providing support via instant messaging, and students accessing library resources from off campus.” This will affect the way language is used as students and teachers begin to use more of these CMC platforms.

At a professional level, it is a common sight for companies to have their computers and laptops hooked up onto the Internet (via wired and wireless Internet connection), and for employees to have individual e-mail accounts. This greatly facilitates internal (among staffs of the company) and external (with other parties outside of one's organization) communication. Mobile communications such as smart phones are increasingly making their way into the corporate world. For instance, in 2008, Apple announced their intention to actively step up their efforts to help companies incorporate the iPhone into their enterprise environment, facilitated by technological developments in streamlining integrated features (push e-mail, calendar and contact management) using ActiveSync.

In general, these new CMCs that are made possible by the Internet have altered the way people use language – there is heightened informality and consequently a growing fear of its deterioration. However, as David Crystal puts it, these should be seen positively as it reflects the power of the creativity of a language.

## **Themes**

The sociolinguistics of the Internet may also be examined through five interconnected themes.

- Multilingualism – It looks at the prevalence and status of various languages on the Internet.
- Language change – From a sociolinguistic perspective, language change is influenced by the physical constraints of technology (e.g. typed text)

and the shifting social-economic priorities such as globalization. It explores the linguistic changes over time, with emphasis on Internet lingo.

- Conversation discourse – It explores the changes in patterns of social interaction and communicative practice on the Internet.
- Stylistic diffusion – It involves the study of the spread of Internet jargons and related linguistic forms into common usage. As language changes, conversation discourse and stylistic diffusion overlap with the aspect of language stylistics.
- *See below: Stylistic perspective*
- Metalanguage and folk linguistics – It involves looking at the way these linguistic forms and changes on the Internet are labelled and discussed (e.g. impact of Internet lingo resulted in the 'death' of the apostrophe and loss of capitalization.)

## **Educational perspective**

The educational perspective of internet linguistics examines the Internet's impact on formal language use, specifically on Standard English, which in turn affects language education. The rise and rapid spread of Internet use has brought about new linguistic features specific only to the Internet platform. These include, but are not limited to, an increase in the use of informal written language, inconsistency in written styles and stylistics and the use of new abbreviations in Internet chats and SMS text messaging, where constraints of technology on word count contributed to the rise of new abbreviations. Such acronyms exist primarily for practical reasons — to reduce the

time and effort required to communicate through these mediums apart from technological limitations. Examples of common acronyms include *lol* (for laughing out loud; a general expression of laughter), *omg* (oh my god) and *gtg* (got to go).

The educational perspective has been considerably established in the research on the Internet's impact on language education. It is an important and crucial aspect as it affects and involves the education of current and future student generations in the appropriate and timely use of informal language that arises from Internet usage. There are concerns for the growing infiltration of informal language use and incorrect word use into academic or formal situations, such as the usage of casual words like "guy" or the choice of the word "preclude" in place of "precede" in academic papers by students. There are also issues with spellings and grammar occurring at a higher frequency among students' academic works as noted by educators, with the use of abbreviations such as "u" for "you" and "2" for "to" being the most common.

Linguists and professors like Eleanor Johnson suspect that widespread mistakes in writing are strongly connected to Internet usage, where educators have similarly reported new kinds of spelling and grammar mistakes in student works. There is, however, no scientific evidence to confirm the proposed connection. Naomi S. Baron (2008) argues in *Always On* that student writings suffer little impact from the use of Internet-mediated communication (IMC) such as internet chat, SMS text messaging and e-mail. A study in 2009 published by the British Journal of Developmental Psychology found that students who regularly texted (sent messages via SMS using a

mobile phone) displayed a wider range of vocabulary and this may lead to a positive impact on their reading development.

Though the use of the Internet resulted in stylistics that are not deemed appropriate in academic and formal language use, Internet use may not hinder language education but instead aid it. The Internet has proven in different ways that it can provide potential benefits in enhancing language learning, especially in second or foreign language learning. Language education through the Internet in relation to Internet linguistics is, most significantly, applied through the communication aspect (use of e-mails, discussion forums, chat messengers, blogs, etc.). IMC allows for greater interaction between language learners and native speakers of the language, providing for greater error corrections and better learning opportunities of standard language, in the process allowing the picking up of specific skills such as negotiation and persuasion.

## **Stylistic perspective**

This perspective examines how the Internet and its related technologies have encouraged new and different forms of creativity in language, especially in literature. It looks at the Internet as a medium through which new language phenomena have arisen. This new mode of language is interesting to study because it is an amalgam of both spoken and written languages. For example, traditional writing is static compared to the dynamic nature of the new language on the Internet where words can appear in different colors and font sizes on the computer screen. Yet, this new mode of language also contains other elements not found in natural languages. One



example is the concept of framing found in e-mails and discussion forums. In replying to e-mails, people generally use the sender's e-mail message as a frame to write their own messages. They can choose to respond to certain parts of an e-mail message while leaving other bits out. In discussion forums, one can start a new thread and anyone regardless of their physical location can respond to the idea or thought that was set down through the Internet. This is something that is usually not found in written language.

Future research also includes new varieties of expressions that the Internet and its various technologies are constantly producing and their effects not only on written languages but also their spoken forms. The communicative style of Internet language is best observed in the CMC channels below, as there are often attempts to overcome technological restraints such as transmission time lags and to re-establish social cues that are often vague in written text.

## **Mobile phones**

Mobile phones (also called cell phones) have an expressive potential beyond their basic communicative functions. This can be seen in text-messaging poetry competitions such as the one held by The Guardian. The 160-character limit imposed by the cell phone has motivated users to exercise their linguistic creativity to overcome them. A similar example of new technology with character constraints is Twitter, which has a 280-character limit. There have been debates as to whether these new abbreviated forms introduced in users' Tweets are "lazy" or whether they are creative fragments of

communication. Despite the ongoing debate, there is no doubt that Twitter has contributed to the linguistic landscape with new lingoes and also brought about a new dimension of communication.

The cell phone has also created a new literary genre – cell phone novels. A typical cell phone novel consists of several chapters which readers download in short installments. These novels are in their "raw" form as they do not go through editing processes like traditional novels. They are written in short sentences, similar to text-messaging. Authors of such novels are also able to receive feedback and new ideas from their readers through e-mails or online feedback channels. Unlike traditional novel writing, readers' ideas sometimes get incorporated into the storyline or authors may also decide to change their story's plot according to the demand and popularity of their novel (typically gauged by the number of download hits). Despite their popularity, there has also been criticism regarding the novels' "lack of diverse vocabulary" and poor grammar.

## **Blogs**

Blogging has brought about new ways of writing diaries and from a linguistic perspective, the language used in blogs is "in its most 'naked' form", published for the world to see without undergoing the formal editing process. This is what makes blogs stand out because almost all other forms of printed language have gone through some form of editing and standardization. David Crystal stated that blogs were "the beginning of a new stage in the evolution of the written

language". Blogs have become so popular that they have expanded beyond written blogs, with the emergence of photoblog, videoblog, audioblog and moblog. These developments in interactive blogging have created new linguistic conventions and styles, with more expected to arise in the future.

## **Virtual worlds**

Virtual worlds provide insights into how users are adapting the usage of natural language for communication within these new mediums. The Internet language that has arisen through user interactions in text-based chatrooms and computer-simulated worlds has led to the development of slangs within digital communities. Examples of these include pwn and noob. Emoticons are further examples of how users have adapted different expressions to suit the limitations of cyberspace communication, one of which is the "loss of emotivity".

Communication in niches such as role-playing games (RPG) of Multi-User domains (MUDs) and virtual worlds is highly interactive, with emphasis on speed, brevity and spontaneity. As a result, CMC is generally more vibrant, volatile, unstructured and open. There are often complex organization of sequences and exchange structures evident in the connection of conversational strands and short turns. Some of the CMC strategies used include capitalization for words such as *EMPHASIS*, usage of symbols such as the asterisk to enclose words as seen in *\*stress\** and the creative use of punctuation like *???!?!?!?*. Symbols are also used for discourse functions, such as the asterisk as a conversational repair marker and

arrows and carats as deixis and referent markers. Besides contributing to these new forms in language, virtual worlds are also being used to teach languages. Virtual world language learning provides students with simulations of real-life environments, allowing them to find creative ways to improve their language skills. Virtual worlds are good tools for language learning among the younger learners because they already see such places as a "natural place to learn and play".

## **E-mail**

One of the most popular Internet-related technologies to be studied under this perspective is e-mail, which has expanded the stylistics of languages in many ways. A study done on the linguistic profile of e-mails has shown that there is a hybrid of speech and writing styles in terms of format, grammar and style. E-mail is rapidly replacing traditional letter-writing because of its convenience, speed and spontaneity. It is often related to informality as it feels temporary and can be deleted easily. However, as this medium of communication matures, e-mail is no longer confined to sending informal messages between friends and relatives. Instead, business correspondences are increasingly being carried out through e-mails. Job seekers are also using e-mails to send their resumes to potential employers. The result of a move towards more formal usages will be a medium representing a range of formal and informal stylistics.

While e-mail has been blamed for students' increased usage of informal language in their written work, David Crystal argues that e-mail is "not a threat, for language education" because e-

mail with its array of stylistic expressiveness can act as a domain for language learners to make their own linguistic choices responsibly. Furthermore, the younger generation's high propensity for using e-mail may improve their writing and communication skills because of the efforts they are making to formulate their thoughts and ideas, albeit through a digital medium.

## **Instant messaging**

Like other forms of online communication, instant messaging has also developed its own acronyms and short forms. However, instant messaging is quite different from e-mail and chat groups because it allows participants to interact with one another in real-time while conversing in private. With instant messaging, there is an added dimension of familiarity among participants. This increased degree of intimacy allows greater informality in language and "typographical idiosyncrasies". There are also greater occurrences of stylistic variation because there can be a very wide age gap between participants. For example, a granddaughter can catch up with her grandmother through instant messaging. Unlike chat groups where participants come together with shared interests, there is no pressure to conform in language here.

## **Applied perspective**

The applied perspective views the linguistic exploitation of the Internet in terms of its communicative capabilities – the good and the bad. The Internet provides a platform where users can experience multilingualism. Although English is still the

dominant language used on the Internet, other languages are gradually increasing in their number of users. The Global Internet usage page provides some information on the number of users of the Internet by language, nationality and geography. This multilingual environment continues to increase in diversity as more language communities become connected to the Internet. The Internet is thus a platform where minority and endangered languages can seek to revive their language use and/or create awareness. This can be seen in two instances where it provides these languages opportunities for progress in two important regards – language documentation and language revitalization.

## **Language documentation**

Firstly, the Internet facilitates language documentation. Digital archives of media such as audio and video recordings not only help to preserve language documentation, but also allows for global dissemination through the Internet. Publicity about endangered languages, such as Webster (2003) has helped to spur a worldwide interest in linguistic documentation.

Foundations such as the Hans Rausing Endangered Languages Project (HRELP), funded by Arcadia also help to develop the interest in linguistic documentation. The HRELP is a project that seeks to document endangered languages, preserve and disseminate documentation materials among others. The materials gathered are made available online under its Endangered Languages Archive (ELAR) program.

Other online materials that support language documentation include the Language Archive Newsletter which provides news and articles about topics in endangered languages. The web version of Ethnologue also provides brief information of all of the world's known living languages. By making resources and information of endangered languages and language documentation available on the Internet, it allows researchers to build on these materials and hence preserve endangered languages.

## **Language revitalization**

Secondly, the Internet facilitates language revitalization. Throughout the years, the digital environment has developed in various sophisticated ways that allow for virtual contact. From e-mails, chats to instant messaging, these virtual environments have helped to bridge the spatial distance between communicators. The use of e-mails has been adopted in language courses to encourage students to communicate in various styles such as conference-type formats and also to generate discussions. Similarly, the use of e-mails facilitates language revitalization in the sense that speakers of a minority language who moved to a location where their native language is not being spoken can take advantage of the Internet to communicate with their family and friends, thus maintaining the use of their native language. With the development and increasing use of telephone broadband communication such as Skype, language revitalization through the internet is no longer restricted to literate users.

Hawaiian educators have been taking advantage of the Internet in their language revitalization programs. The graphical bulletin board system, Leoki (Powerful Voice), was established in 1994. The content, interface and menus of the system are entirely in the Hawaiian language. It is installed throughout the immersion school system and includes components for e-mails, chat, dictionary and online newspaper among others. In higher institutions such as colleges and universities where the Leoki system is not yet installed, the educators make use of other software and Internet tools such as Daedalus Interchange, e-mails and the Web to connect students of Hawaiian language with the broader community.

Another use of the Internet includes having students of minority languages write about their native cultures in their native languages for distant audiences. Also, in an attempt to preserve their language and culture, Occitan speakers have been taking advantage of the Internet to reach out to other Occitan speakers from around the world. These methods provide reasons for using the minority languages by communicating in it. In addition, the use of digital technologies, which the young generation think of as 'cool', will appeal to them and in turn maintain their interest and usage of their native languages.

## **Exploitation of the Internet**

The Internet can also be exploited for activities such as terrorism, internet fraud and pedophilia. In recent years, there has been an increase in crimes that involved the use of the Internet such as e-mails and Internet Relay Chat (IRC), as it is



relatively easy to remain anonymous. These conspiracies carry concerns for security and protection. From a forensic linguistic point of view, there are many potential areas to explore. While developing a chat room child protection procedure based on search terms filtering is effective, there is still minimal linguistically orientated literature to facilitate the task. In other areas, it is observed that the Semantic Web has been involved in tasks such as personal data protection, which helps to prevent fraud.

## **Dimensions**

The dimensions covered in this section include looking at the Web as a corpus and issues of language identification and normalization. The impacts of internet linguistics on everyday life are examined under the spread and influence of Internet stylistics, trends of language change on the Internet and conversation discourse.

### **The Web as a corpus**

With the Web being a huge reservoir of data and resources, language scientists and technologists are increasingly turning to the web for language data. Corpora were first formally mentioned in the field of computational linguistics at the 1989 ACL meeting in Vancouver. It was met with much controversy as they lacked theoretical integrity leading to much skepticism of their role in the field, until the publication of the journal 'Using Large Corpora' in 1993 that the relationship between computational linguistics and corpora became widely accepted.

To establish whether the Web is a corpus, it is worthwhile to turn to the definition established by McEnery and Wilson (1996, pp 21).

In principle, any collection of more than one text can be called a corpus. . . . But the term “corpus” when used in the context of modern linguistics tends most frequently to have more specific connotations than this simple definition provides for. These may be considered under four main headings: sampling and representativeness, finite size, machine-readable form, a standard reference.

- — Tony McEnery and Andrew Wilson, *Corpus Linguistics*

Relating closer to the Web as a Corpus, Manning and Schütze (1999, pp 120) further streamlines the definition:

In Statistical NLP [Natural Language Processing], one commonly receives as a corpus a certain amount of data from a certain domain of interest, without having any say in how it is constructed. In such cases, having more training data is normally more useful than any concerns of balance, and one should simply use all the text that is available.

- — Christopher Manning and Hinrich Schütze, *Foundations of Statistical Language Processing*

Hit counts were used for carefully constructed search engine queries to identify rank orders for word sense frequencies, as an input to a word sense disambiguation engine. This method was further explored with the introduction of the concept of a parallel corpora where the existing Web pages that exist in

parallel in local and major languages be brought together. It was demonstrated that it is possible to build a language-specific corpus from a single document in that specific language.

## **Themes**

There has been much discussion about the possible developments in the arena of the Web as a corpus. The development of using the web as a data source for word sense disambiguation was brought forward in The EU MEANING project in 2002. It used the assumption that within a domain, words often have a single meaning, and that domains are identifiable on the Web. This was further explored by using Web technology to gather manual word sense annotations on the Word Expert Web site.

In areas of language modeling, the Web has been used to address data sparseness. Lexical statistics have been gathered for resolving prepositional phrase attachments, while Web document were used to seek a balance in the corpus.

In areas of information retrieval, a Web track was integrated as a component in the community's TREC evaluation initiative. The sample of the Web used for this exercise amount to around 100GB, comprising of largely documents in the .gov top level domain.

## **British National Corpus**

The British National Corpus contains ample information on the dominant meanings and usage patterns for the 10,000 words that forms the core of English.

The number of words in the British National Corpus (ca 100 million) is sufficient for many empirical strategies for learning about language for linguists and lexicographers, and is satisfactory for technologies that utilize quantitative information about the behavior of words as input (parsing).

However, for some other purposes, it is insufficient, as an outcome of the Zipfian nature of word frequencies. Because the bulk of the lexical stock occurs less than 50 times in the British National Corpus, it is insufficient for statistically stable conclusions about such words. Furthermore, for some rarer words, rare meanings of common words, and combinations of words, no data has been found. Researchers find that probabilistic models of language based on very large quantities of data are better than ones based on estimates from smaller, cleaner data sets.

## **The multilingual Web**

The Web is clearly a multilingual corpus. It is estimated that 71% of the pages (453 million out of 634 million Web pages indexed by the Excite engine) were written in English, followed by Japanese (6.8%), German (5.1%), French (1.8%), Chinese (1.5%), Spanish (1.1%), Italian (0.9%), and Swedish (0.7%).

A test to find contiguous words like ‘deep breath’ revealed 868,631 Web pages containing the terms in AlltheWeb. The number found through the search engines are more than three times the counts generated by the British National Corpus, indicating the significant size of the English corpus available on the Web.

The massive size of text available on the Web can be seen in the analysis of controlled data in which corpora of different languages were mixed in various proportions. The estimated Web size in words by AltaVista saw English at the top of the list with 76,598,718,000 words. The next is German, with 7,035,850,000 words along with 6 other languages with over a billion hits. Even languages with fewer hits on the Web such as Slovenian, Croatian, Malay, and Turkish have more than one hundred million words on the Web. This reveals the potential strength and accuracy of using the Web as a Corpus given its significant size, which warrants much additional research such as the project currently being carried out by the British National Corpus to exploit its scale.

## **Challenges**

In areas of language modeling, there are limitations on the applicability of any language model as the statistics for different types of text will be different. When a language technology application is put into use (applied to a new text type), it is not certain that the language model will fare in the same way as how it would when applied to the training corpus. It is found that there are substantial variations in model performance when the training corpus changes. This lack of

theory types limits the assessment of the usefulness of language-modeling work.

As Web texts are easily produced (in terms of cost and time) and with many different authors working on them, it often results in little concern for accuracy. Grammatical and typographical errors are regarded as “erroneous” forms that cause the Web to be a dirty corpus. Nonetheless, it may still be useful even with some noise.

The issue of whether sublanguages should be included remains unsettled. Proponents of it argue that with all sublanguages removed, it will result in an impoverished view of language. Since language is made up of lexicons, grammar and a wide array of different sublanguages, they should be included. However, it is not until recently that it became a viable option. Striking a middle ground by including some sublanguages is contentious because it's an arbitrary issue of which to include and which not.

The decision of what to include in a corpus lies with corpus developers, and it has been done so with pragmatism. The desiderata and criteria used for the British National Corpus serves as a good model for a general-purpose, general-language corpus with the focus of being representative replaced with being balanced.

Search engines such as Google serves as a default means of access to the Web and its wide array of linguistics resources. However, for linguists working in the field of corpora, there presents a number of challenges. This includes the limited instances that are presented by the search engines (1,000 or 5,000 maximum); insufficient context for each instance (Google

provides a fragment of around ten words); results selected according to criteria that are distorted (from a linguistic point of view) as search term in titles and headings often occupy the top results slots; inability to allow searches to be specified according to linguistic criteria, such as the citation form for a word, or word class; unreliability of statistics, with results varying according to search engine load and many other factors. At present, in view of the conflicts of priorities among the different stakeholders, the best solution is for linguists to attempt to correct these problems by themselves. This will then lead to a large number of possibilities opening in the area of harnessing the rich potential of the Web.

## **Representation**

Despite the sheer size of the Web, it may still not be representative of all the languages and domains in the world, and neither are other corpora. However, the huge quantities of text, in numerous languages and language types on a huge range of topics makes it a good starting point that opens up to a large number of possibilities in the study of corpora.

## **Impact of its spread and influence**

Stylistics arising from Internet usage has spread beyond the new media into other areas and platforms, including but not limited to, films, music and literary works. The infiltration of Internet stylistics is important as mass audiences are exposed to the works, reinforcing certain Internet specific language styles which may not be acceptable in standard or more formal forms of language.

Apart from internet slang, grammatical errors and typographical errors are features of writing on the Internet and other CMC channels. As users of the Internet gets accustomed to these errors, it progressively infiltrates into everyday language use, in both written and spoken forms. It is also common to witness such errors in mass media works, from typographical errors in news articles to grammatical errors in advertisements and even internet slang in drama dialogues.

The more the internet is incorporated into daily life, the greater the impact it has on formal language. This is especially true in modern Language Arts classes through the use of smart phones, tablets, and social media. Students are exposed to the language of the internet more than ever, and as such, the grammatical structure and slang of the internet are bleeding into their formal writing. Full immersion into a language is always the best way to learn it. Mark Lester in his book *Teaching Grammar and Usage* states, "The biggest single problem that basic writers have in developing successful strategies for coping with errors is simply their lack of exposure to formal written English...We would think it absurd to expect a student to master a foreign language without extensive exposure to it." Since students are immersed in internet language, that is the form and structure they are mirroring.

In addition, the rise of the Internet and overall immersion of people within it has brought forth a new wave over internet activism that has an impact on the public each and every day.



## **Memes**

The origin of the term 'meme' can be traced back to Richard Dawkins, an ethologist, where he describes it as "a noun that conveys the idea of a unit of cultural transmission, or a unite of imitation." The term was later adapted to the realm of the Internet by David Beskow, Sumeet Kumar, and Kathleen Carley, wherein they labeled Internet memes as "any digital unit that transfers culture."

## **Mass media**

There has been instances of television advertisements using Internet slang, reinforcing the penetration of Internet stylistics in everyday language use. For example, in the Cingular commercial in the United States, acronyms such as "BFF Jill" (which means "Best Friend Forever, Jill") were used. More businesses have adopted the use of Internet slang in their advertisements as the more people are growing up using the Internet and other CMC platforms, in an attempt to relate and connect to them better. Such commercials have received relatively enthusiastic feedback from its audiences.

The use of Internet lingo has also spread into the arena of music, significantly seen in popular music. A recent example is Trey Songz's lyrics for "LOL :-)", which incorporated many Internet lingo and mentions of Twitter and texting.

The spread of Internet linguistics is also present in films made by both commercial and independent filmmakers. Though

primarily screened at film festivals, DVDs of independent films are often available for purchase over the internet including paid-live-streamings, making access to films more easily available for the public. The very nature of commercial films being screened at public cinemas allows for the wide exposure to the mainstream mass audience, resulting in a faster and wider spread of Internet slangs. The latest commercial film is titled "LOL" (acronym for *Laugh Out Loud* or *Laughing Out Loud*), starring Miley Cyrus and Demi Moore. This movie is a 2011 remake of the Lisa Azuelos' 2008 popular French film similarly titled "LOL (Laughing Out Loud)".

The use of internet slangs is not limited to the English language but extends to other languages as well. The Korean language has incorporated the English alphabet in the formation of its slang, while others were formed from common misspellings arising from fast typing. The new Korean slang is further reinforced and brought into everyday language use by television shows such as soap operas or comedy dramas like "High Kick Through the Roof" released in 2009.

## **Linguistic future of the Internet**

With the emergence of greater computer/Internet mediated communication systems, coupled with the readiness with which people adapt to meet the new demands of a more technologically sophisticated world, it is expected that users will continue to remain under pressure to alter their language use to suit the new dimensions of communication.

As the number of Internet users increase rapidly around the world, the cultural background, linguistic habits and language differences among users are brought into the Web at a much faster pace. These individual differences among Internet users are predicted to significantly impact the future of Internet linguistics, notably in the aspect of the multilingual web. As seen from 2000 to 2010, Internet penetration has experienced its greatest growth in non-English speaking countries such as China and India and countries in Africa, resulting in more languages apart from English penetrating the Web.

Also, the interaction between English and other languages is predicted to be an important area of study. As global users interact with each other, possible references to different languages may continue to increase, resulting in formation of new Internet stylistics that spans across languages. Chinese and Korean languages have already experienced English language's infiltration leading to the formation of their multilingual Internet lingo.

At current state, the Internet provides a form of education and promotion for minority languages. However, similar to how cross-language interaction has resulted in English language's infiltration into Chinese and Korean languages to form new slangs, minority languages are also affected by the more common languages used on the Internet (such as English and Spanish). While language interaction can cause a loss in the authentic standard of minority languages, familiarity of the majority language can also affect the minority languages in adverse ways. For example, users attempting to learn the minority language may opt to read and understand about it in a majority language and stop there, resulting in a loss instead

of gain in the potential speakers of the minority language. Also, speakers of minority languages may be encouraged to learn the more common languages that are being used on the Web in order to gain access to more resources, and in turn leading to a decline in their usage of their own language. The future of endangered minority languages in view of the spread of Internet remains to be observed.

## **Stylistics**

**Stylistics**, a branch of applied linguistics, is the study and interpretation of texts of all types and/or spoken language in regard to their linguistic and tonal style, where style is the particular variety of language used by different individuals and/or in different situations or settings. For example, the vernacular, or everyday language may be used among casual friends, whereas more formal language, with respect to grammar, pronunciation or accent, and lexicon or choice of words, is often used in a cover letter and résumé and while speaking during a job interview.

As a discipline, stylistics links literary criticism to linguistics. It does not function as an autonomous domain on its own, and it can be applied to an understanding of literature and journalism as well as linguistics. Sources of study in stylistics may range from canonical works of writing to popular texts, and from advertising copy to news, non-fiction, and popular culture, as well as to political and religious discourse. Indeed, as recent work in critical stylistics, multimodal stylistics and mediated stylistics has made clear, non-literary texts may be of just as much interest to stylisticians as literary ones.

Literariness, in other words, is here conceived as 'a point on a cline rather than as an absolute'.

Stylistics as a conceptual discipline may attempt to establish principles capable of explaining particular choices made by individuals and social groups in their use of language, such as in the literary production and reception of genre, the study of folk art, in the study of spoken dialects and registers, and can be applied to areas such as discourse analysis as well as literary criticism.

Plain language has different features. Common stylistic features are using dialogue, regional accents and individual idioms (or idiolects). Stylistically, also sentence length prevalence and language register use.

## **Early twentieth century**

The analysis of literary style goes back to the study of classical rhetoric, though modern stylistics has its roots in Russian Formalism and the related Prague School of the early twentieth century.

In 1909, Charles Bally proposed stylistics as a distinct academic discipline to complement Saussurean linguistics. For Bally, Saussure's linguistics by itself couldn't fully describe the language of personal expression. Bally's programme fits well with the aims of the Prague School.

Taking forward the ideas of the Russian Formalists, the Prague School built on the concept of *foregrounding*, where it is assumed that poetic language is considered to stand apart

from non-literary background language, by means of *deviation* (from the norms of everyday language) or *parallelism*. According to the Prague School, however, this background language isn't constant, and the relationship between poetic and everyday language is therefore always shifting.

## Late twentieth century

Roman Jakobson had been an active member of the Russian Formalists and the Prague School, before emigrating to America in the 1940s. He brought together Russian Formalism and American New Criticism in his *Closing Statement* at a conference on stylistics at Indiana University in 1958. Published as *Linguistics and Poetics* in 1960, Jakobson's lecture is often credited with being the first coherent formulation of stylistics, and his argument was that the study of poetic language should be a sub-branch of linguistics. The *poetic function* was one of six general functions of language he described in the lecture.

Michael Halliday is an important figure in the development of British stylistics. His 1971 study *Linguistic Function and Literary Style: An Inquiry into the Language of William Golding's **The Inheritors*** is a key essay. One of Halliday's contributions has been the use of the term register to explain the connections between language and its context. For Halliday register is distinct from dialect. Dialect refers to the habitual language of a particular user in a specific geographical or social context. Register describes the choices made by the user, choices which depend on three variables: *field* ("what the participants... are actually engaged in doing", for instance,

discussing a specific subject or topic), *tenor* (who is taking part in the exchange) and *mode* (the use to which the language is being put).

Fowler comments that different fields produce different language, most obviously at the level of vocabulary (Fowler. 1996, 192). The linguist David Crystal points out that Halliday's 'tenor' stands as a roughly equivalent term for 'style', which is a more specific alternative used by linguists to avoid ambiguity (Crystal. 1985, 292). Halliday's third category, *mode*, is what he refers to as the symbolic organisation of the situation. Downes recognises two distinct aspects within the category of mode and suggests that not only does it describe the relation to the medium: written, spoken, and so on, but also describes the genre of the text (Downes. 1998, 316). Halliday refers to genre as pre-coded language, language that has not simply been used before, but that predetermines the selection of textual meanings. The linguist William Downes makes the point that the principal characteristic of register, no matter how peculiar or diverse, is that it is obvious and immediately recognisable (Downes. 1998, 309).

## **Literary stylistics**

In *The Cambridge Encyclopedia of Language*, Crystal observes that, in practice, most stylistic analysis has attempted to deal with the complex and 'valued' language within literature, i.e. 'literary stylistics'. He goes on to say that in such examination the scope is sometimes narrowed to concentrate on the more striking features of literary language, for instance, its 'deviant' and abnormal features, rather than the broader structures that

are found in whole texts or discourses. For example, the compact language of poetry is more likely to reveal the secrets of its construction to the stylistician than is the language of plays and novels (Crystal. 1987, 71).

## **Poetry**

As well as conventional styles of language there are the unconventional – the most obvious of which is poetry. In *Practical Stylistics*, HG Widdowson examines the traditional form of the epitaph, as found on headstones in a cemetery. For example:

- His memory is dear today
- As in the hour he passed away.
- (Ernest C. Draper 'Ern'. Died 4.1.38)
- (Widdowson. 1992, 6)

Widdowson makes the point that such sentiments are usually not very interesting and suggests that they may even be dismissed as 'crude verbal carvings' and crude verbal disturbance (Widdowson, 3). Nevertheless, Widdowson recognises that they are a very real attempt to convey feelings of human loss and preserve affectionate recollections of a beloved friend or family member. However, what may be seen as poetic in this language is not so much in the formulaic phraseology but in where it appears. The verse may be given undue reverence precisely because of the sombre situation in which it is placed. Widdowson suggests that, unlike words set in stone in a graveyard, poetry is unorthodox language that vibrates with inter-textual implications (Widdowson. 1992, 4).



Two problems with a stylistic analysis of poetry are noted by PM Wetherill in *Literary Text: An Examination of Critical Methods*. The first is that there may be an over-preoccupation with one particular feature that may well minimise the significance of others that are equally important (Wetherill. 1974, 133). The second is that any attempt to see a text as simply a collection of stylistic elements will tend to ignore other ways whereby meaning is produced (Wetherill. 1974, 133).

## **Implicature**

In 'Poetic Effects' from *Literary Pragmatics*, the linguist Adrian Pilkington analyses the idea of 'implicature', as instigated in the previous work of Dan Sperber and Deirdre Wilson. Implicature may be divided into two categories: 'strong' and 'weak' implicature, yet between the two extremes there are a variety of other alternatives. The strongest implicature is what is emphatically implied by the speaker or writer, while weaker implicatures are the wider possibilities of meaning that the hearer or reader may conclude.

Pilkington's 'poetic effects', as he terms the concept, are those that achieve most relevance through a wide array of weak implicatures and not those meanings that are simply 'read in' by the hearer or reader. Yet the distinguishing instant at which weak implicatures and the hearer or reader's conjecture of meaning diverge remains highly subjective. As Pilkington says: 'there is no clear cut-off point between assumptions which the speaker certainly endorses and assumptions derived purely on the hearer's responsibility.' (Pilkington. 1991, 53) In addition, the stylistic qualities of poetry can be seen as an

accompaniment to Pilkington's poetic effects in understanding a poem's meaning.

## **Tense**

Widdowson points out that in Samuel Taylor Coleridge's poem "The Rime of the Ancient Mariner" (1798), the mystery of the Mariner's abrupt appearance is sustained by an idiosyncratic use of tense. (Widdowson. 1992, 40) For instance, the Mariner 'holds' the wedding-guest with his 'skinny hand' in the present tense, but releases it in the past tense ('...his hands dropt he. '); only to hold him again, this time with his 'glittering eye', in the present (Widdowson. 1992, 41).

## **The point of poetry**

Widdowson notices that when the content of poetry is summarised, it often refers to very general and unimpressive observations, such as 'nature is beautiful; love is great; life is lonely; time passes', and so on (Widdowson. 1992, 9). But to say:

- Like as the waves make towards the pebbled shore,
- So do our minutes hasten to their end ...
- William Shakespeare, '60'.

Or, indeed:

- Love, all alike, no season knows nor clime,
- Nor hours, days months, which are the rags of time
- ...
- John Donne, 'The Sun Rising', *Poems* (1633)

This language gives the reader a new perspective on familiar themes and allows us to look at them without the personal or social conditioning that we unconsciously associate with them (Widdowson. 1992, 9). So, although the reader may still use the same exhausted words and vague terms like 'love', 'heart' and 'soul' to refer to human experience, to place these words in a new and refreshing context allows the poet the ability to represent humanity and communicate honestly. This, in part, is stylistics, and this, according to Widdowson, is the point of poetry (Widdowson. 1992, 76).

## Chapter 5

# Forensic Linguistics

**Forensic linguistics, legal linguistics, or language and the law**, is the application of linguistic knowledge, methods, and insights to the forensic context of law, language, crime investigation, trial, and judicial procedure. It is a branch of applied linguistics.

There are principally three areas of application for linguists working in forensic contexts:

- understanding language of the written law,
- understanding language use in forensic and judicial processes, and
- the provision of linguistic evidence.

The discipline of forensic linguistics is not homogeneous; it involves a range of experts and researchers in different areas of the field.

## History

The phrase *forensic linguistics* first appeared in 1968 when Jan Svartvik, a professor of linguistics, used it in an analysis of statements by Timothy John Evans. It was in regard to re-analyzing the statements given to police at Notting Hill police station, England, in 1949 in the case of an alleged murder by Evans. Evans was suspected of murdering his wife and baby, and was tried and hanged for the crime. Yet, when Svartvik

studied the statements allegedly given by Evans, he found that there were different stylistic markers involved, and Evans did not actually give the statements to the police officers as had been stated at the trial. Sparked by this case, early forensic linguistics in the UK were focused on questioning the validity of police interrogations. As seen in numerous famous cases (e.g. the convictions of Derek Bentley, the Guildford Four, the Bridgewater Three, etc.), many of the major concerns were of the statements police officers used. Numerous times, the topic of *police register* came up – the type of stylist language and vocabulary used by officers of the law when transcribing witness statements.

During the early days of forensic linguistics in the United Kingdom, the legal defense for many criminal cases questioned the authenticity of police statements. At the time, customary police procedure for taking suspects' statements dictated that it be in a specific format, rather than in the suspect's own words. Statements by witnesses are very seldom made in a coherent or orderly fashion, with speculation and backtracking done out loud. The delivery is often too fast-paced, causing important details to be left out.

In the US, forensic linguistics can be traced back as early as 1927 to a ransom note in Corning, New York. As the Associated Press reported in the article "Think Corning Girl Wrote Ransom Note", "Duncan McLure, of Johnson City, uncle of the [kidnapped] girl, is the only member of the family to spell his name 'McLure' instead of 'McClure.' The letter he received, supposedly from the kidnappers, was addressed to him by the proper name, indicating that the writer was familiar with the difference in spelling." Other work of forensic linguistics in the

United States concerned the rights of individuals with regard to understanding their Miranda rights during the interrogation process. The 1963 case of Ernesto Miranda was pivotal to the beginnings of the forensic linguistics field. His case led to the creation of the Miranda Rights and pushed the focus of forensic linguistics on witness questioning rather than police statements. Various cases came about that challenged whether or not suspects truly understood what their rights meant – leading to a distinction of coercive versus voluntary interrogations. An early application of forensic linguistics in the United States was related to the status of trademarks as words or phrases in the language. One of the bigger cases involved fast food giant McDonald's claiming that it had originated the process of attaching unprotected words to the "Mc" prefix (referred to as McWords) and was unhappy with Quality Inns International's intention of opening a chain of economy hotels to be called "McSleep".

In the 1980s, Australian linguists discussed the application of linguistics and sociolinguistics to legal issues. They discovered that a phrase such as "the same language" is open to interpretation. Aboriginal people have their own understanding and use of "English", something that is not always appreciated by speakers of the dominant version of English, i.e. "white English". The Aboriginal people also bring their own culturally-based interactional styles to the interview.

The 2000s saw a considerable shift in the field of forensic linguistics, which has been described as a coming-of-age of the discipline, as it spread to many countries around the world, from Europe to Australia and Japan. Today, not only does the field have professional associations such as the International

Association of Forensic Linguistics (IAFL), founded in 1993, and the Austrian Association for Legal Linguistics (AALL), founded in 2017, but it can now provide the scientific community with a range of textbooks such as Coulthard and Johnson (2007), Gibbons (2003), and Olsson (2008).

## **Areas of study**

The range of topics within forensic linguistics is diverse, but research occurs in the following areas:

### **The language of legal texts**

The study of the language of legal texts encompasses a wide range of forensic texts. That includes the study of text types and forms of analysis. Any text or item of spoken language can potentially be a forensic text when it is used in a legal or criminal context. This includes analyzing the linguistics of documents as diverse as Acts of Parliament (or other law-making bodies), private wills, court judgments and summonses, and the statutes of other bodies, such as States and government departments. One important area is that of the transformative effect of Norman French and Ecclesiastic Latin on the development of the English common law, and the evolution of the legal specifics associated with it. It can also refer to the ongoing attempts at making legal language more comprehensible to laypeople.

One part of the language of legal texts encompasses the Miranda warning in the United States. These warnings let the defendant know that they have the right to be silent since

whatever they say from the point they are in police custody can and will be used against them in a court of law. The recipients who are advised of these rights must have a certain level of competency in the English language in order to completely understand the warning.

## **The language of legal processes**

Among other things, this area examines language as it is used in cross-examination, evidence presentation, judge's direction, police cautions, police testimonies in court, summing up to a jury, interview techniques, the questioning process in court, and in other areas such as police interviews.

Police officers use specific language to elicit certain responses from civilians. Because of a police officer's social stature, and the way they often phrase "requests" as "commands", people may be confused as to what their rights are when they are being questioned by police. Officers use linguistic tactics including putting the blame onto the victim and asking questions with ambiguous phrasing to elicit specific responses from people (Solan, L. & Tiersma, P, 2005).

When a victim is invoking their right to a lawyer, there are directions stating that the request may not come off as ambiguous. In fact, if the request is not stated in a way that the officer deems to be clear, the victim may not receive their request for counsel at all.

During the examination process, language plays a substantial role in the presentation of a story to the courtroom. A defendant's ambiguity may be deemed unacceptable. The



language used by the lawyer to construct the story to the courtroom elicits specific responses from the witness, and specific emotions from the jury. For example, in an instance where a lawyer is examining a hostile witness, they will often use language to limit the response of the witness, in order to avoid having the witness present conflicting evidence. In this instance, yes/no questions will be targeted, and questions with room for elaboration, such as wh-formation questions, will likely be avoided. In a situation where a lawyer interviews a friendly witness whose testimony could potentially strengthen the story constructed by the lawyer, the opposite may occur, where wh-questions are targeted to allow for elaboration (Olsson, Luchjenborers, 2013).

Lawyers employ specific tactics for both themselves and their witnesses to come off as more or less truthful to the jury and the people of the courtroom. For example, the lawyer may refer to the witness by their first name or a nickname to humanize the witness, or they may speak using slang in order to create less social distance between themselves and the courtroom. The lawyer may also avoid using slang, and instead use complicated law terminology to set themselves apart from the courtroom and define their status (Olsson, Luchjenborers, 2013).

The lawyer works in constructing the language of the legal process of the courtroom, and specific witnesses may respond to the lawyer's questions in different ways, eliciting new language tactics and opinions from the jury. For example, witnesses may use direct or indirect speech based on their previous societal experiences, gender differences, socioeconomic differences, or differences in education level.

Using particular dialects, slang, or sentence formations could assist in making the witness more or less truthful to the jury (Olsson, Luchjenborers, 2013).

Since language can be used to elicit responses in the legal process, the right to an interpreter plays a part in the fairness of the trial. "The right to an interpreter is essentially a procedural right that derives from the right to a fair trial: everyone charged with a criminal offence has the right to certain minimum procedural guarantees, and these include the right to the free assistance of an interpreter where s/he cannot understand or speak the language of the concerned court."

## **Forensic text types**

### **Emergency call**

In an emergency call, the recipient or emergency operator's ability to extract primarily linguistic information in threatening situations and to come up with the required response in a timely manner is crucial to the successful completion of the call. Emphasis on intonation, voice pitch, and the extent to which there is cooperation between the caller and the recipient at any one time are also very important in analyzing an emergency call. Full cooperation includes frank and timely responses.

Urgency plays a role in emergency calls, so hesitations, signs of evasiveness, and incomplete or overly short answers indicate that the caller might be making a false or hoax call. A genuine call has distinctive interlocking and slight overlap of turns.

The recipient trusts the caller to provide accurate information and the caller trusts the recipient to ask only pertinent questions. If the caller uses a rising pitch at the end of every turn, it might represent a lack of commitment; the recipient's use of a rising pitch indicates doubt or desire for clarification. The call ideally moves from nil knowledge on the part of the recipient to a maximum amount of knowledge in a minimum possible period of time. This makes the emergency call unlike any other kind of service encounter.

## **Ransom demands or other threat communication**

Threat is a counterpart of a promise and is an important feature in a ransom demand. Ransom demands are also examined to identify between genuine and false threats. An example of a ransom note analysis can be seen in the case of the Lindbergh kidnapping, where the first ransom note (sometimes referred to as the Nursery Note) stated: "We warn you for making anyding public or for notify the Polise the child is in gut care" [sic]. In the sentence, the kidnapper makes the claim that the child is in good hands, but to make such a claim, the note would have to have been written before the perpetrator entered the premises. Therefore, the claim is false (at the time of writing) since the kidnapper had not even encountered the child when he wrote the note.

Kidnappers may write statements that later end up being true, such as "your child is being held in a private location" being written ahead of time.

Ransom demands in the style of written notes have been present in many notable cases. The style of writing used in a ransom note is examined by forensic linguists in order to determine the writing's true intent, as well as to determine who wrote the note. Forensic linguists look at factors such as syntactic structures, stylistic patterns, punctuation, and even spelling while analyzing ransom notes. In the case of the Lindbergh ransom note, forensic linguists compared similarities of writing styles from the note to that of writing of the suspect, creating a better chance at discovering who wrote the note.

## **Suicide letters**

A suicide note is typically brief, concise, and highly propositional with a degree of evasiveness. A credible suicide letter must be making a definite unequivocal proposition in a situational context. The proposition of genuine suicide is thematic, directed to the addressee (or addressees), and relevant to the relationship between them and the writer. Suicide notes generally have sentences alluding to the act of killing oneself, or the method of suicide that was undertaken. The contents of a suicide note could be intended to make the addressee suffer or feel guilt. Genuine suicide letters are short, typically less than 300 words in length. Extraneous or irrelevant material is often excluded from the text.

## **Death row statements**

Death row statements either admit the crime, leaving the witness with an impression of honesty and forthrightness, or

deny the crime, leaving the witness with an impression of innocence. They may also denounce witnesses as dishonest, critique law enforcement as corrupt in an attempt to portray innocence, or seek an element of revenge in their last moments (Olsson 2004). Death row statements are made within the heavily institutionalized setting of death row prisons.

The Forensic Linguistics Institute holds a corpus of these documents and is conducting research on them.

## **Social media**

Social media statements are often context-specific, and their interpretation can be highly subjective. Forensic application of a selection of stylistic and stylometric techniques has been done in a simulated authorship attribution case involving texts in relation to Facebook. Analysis of social media postings can reveal whether they are illegal (e.g. sex trade) or unethical (e.g. intended to harm) or whether they are not (e.g. simply provocative or free speech).

## **Use of linguistic evidence in legal proceedings**

These areas of application have varying degrees of acceptability or reliability within the field. Linguists have provided evidence in:

- Trademark and other intellectual property disputes
- Disputes of meaning and use

- Author identification (determining who wrote an anonymous text by making comparisons to known writing samples of a suspect; such as threat letters, mobile phone texts or emails)
- Forensic stylistics (identifying cases of plagiarism)
- Voice identification, also known as forensic phonetics (used to determine, through acoustic qualities, if the voice on a tape recorder is that of the defendant)
- Discourse analysis (the analysis of the structure of written or spoken utterance to determine who is introducing topics or whether a suspect is agreeing to engage in criminal conspiracy)
- Language analysis (forensic dialectology) tracing the linguistic history of asylum seekers (Language Analysis for the Determination of Origin)
- Reconstruction of mobile phone text conversations
- Forensic phonetics

Specialist databases of samples of spoken and written natural language (called corpora) are now frequently used by forensic linguists. These include corpora of suicide notes, mobile phone texts, police statements, police interview records, and witness statements. They are used to analyze language, understand how it is used, and to reduce the effort needed to identify words that tend to occur near each other (collocations or collocates).

## **Author identification**

The identification of whether a given individual said or wrote something relies on analysis of their idiolect, or particular

patterns of language use (vocabulary, collocations, pronunciation, spelling, grammar, etc.). The idiolect is a theoretical construct based on the idea that there is linguistic variation at the group level and hence there may also be linguistic variation at the individual level. William Labov has stated that nobody has found a "homogenous data" in idiolects, and there are many reasons why it is difficult to provide such evidence.

Firstly, language is not an inherited property, but one which is socially acquired. Because acquisition is continuous and life-long, an individual's use of language is always susceptible to variation from a variety of sources, including other speakers, the media, and macro-social changes. Education can have a profoundly homogenizing effect on language use. Research into authorship identification is ongoing. The term authorship attribution is now felt to be too deterministic.

The paucity of documents (ransom notes, threatening letters, etc.) in most criminal cases in a forensic setting means there is often too little text upon which to base a reliable identification. However, the information provided may be adequate to eliminate a suspect as an author or narrow down an author from a small group of suspects.

Authorship measures that analysts use include word length average, average number of syllables per word, article frequency, type-token ratio, punctuation (both in terms of overall density and syntactic boundaries), and the measurements of hapax legomena (unique words in a text). Statistical approaches include factor analysis, Bayesian

statistics, Poisson distribution, multivariate analysis, and discriminant function analysis of function words.

The CUSUM (cumulative sum) method for text analysis has also been developed. CUSUM analysis works even on short texts and relies on the assumption that each speaker has a unique set of habits, thus rendering no significant difference between their speech and writing. Speakers tend to utilize two- to three-letter words in a sentence and their utterances tend to include vowel-initial words.

In order to carry out the CUSUM test on habits of utilizing two- to three-letter words and vowel-initial words in a sentential clause, the occurrences of each type of word in the text must be identified and the distribution plotted in each sentence. The CUSUM distribution for these two habits will be compared with the average sentence length of the text. The two sets of values should track each other. Any altered section of the text would show a distinct discrepancy between the values of the two reference points. The tampered-with section will exhibit a different pattern from the rest of the text.

## **Forensic stylistics**

This discipline subjects written or spoken materials (or both) to scientific analysis for determination and measurement of content and meaning, speaker identification, or determination of authorship in identifying plagiarism.

One of the earliest cases where forensic stylistics was used to detect plagiarism was the case of Helen Keller's short story "The Frost King", in which the deaf-blind American author was



accused of plagiarism in 1892. An investigation revealed that "The Frost King" had been plagiarized from Margaret Canby's "Frost Fairies", which had been read to Keller some time earlier. Keller was found to have made only minute changes to common words and phrases and used less common words to say the same thing, suggesting mere alterations to original ideas. Keller used "vast wealth" instead of "treasure" (230 times less common in the language), "bethought" instead of "concluded" (approximately 450 times less common), and "bade them" instead of "told them" (approximately 30 times less common). Keller used the phrase "ever since that time", but Canby chose "from that time" (the latter 50 times more common than the former). Keller also used "I cannot imagine", but Canby used "I do not know". "Know" is approximately ten times more common than "imagine". Keller relied on a lexis that is less common than that of Canby. The Flesch and Flesch-Kincaid readability test showed that Canby's text had more originality than Keller's. Canby's text obtained a higher grade on the reading ease scale compared to Keller's. The distinctions between Keller and Canby's text are at the lexical and phrasal level.

Other examples of plagiarism include the cases between Richard Condon, author of *The Manchurian Candidate*, and English novelist Robert Graves and between Martin Luther King Jr. and Archibald Carey. Judging by the text in *The Manchurian Candidate*, Condon's work is rich in clichés such as "in his superstitious heart of hearts". While Helen Keller took pride in using rare phrases and avoided common source words, Condon was fond of expanding existing words into phrases and existing phrases into more extensive ones. Condon was also found to have borrowed from a wide range of Graves'

work. However, in the plagiarism case between King and Archibald Carey, almost half of King's doctoral dissertation was discovered to have been copied from another theology student. King simply changed the names of the mountains and used much more alliteration and assonance. Carey's and Graves' texts (source texts) were noticeably shorter, pithier, and simpler in structure, while Condon's and King's texts relied on "purple" devices, extending the existing text and flourishing their language significantly.

Another famous example is that in the case of Theodore Kaczynski, who was eventually convicted of being the "Unabomber". Family members recognized his writing style in the published 35,000-word *Industrial Society and Its Future* (commonly called the "Unabomber Manifesto") and notified the authorities. FBI agents searching Kaczynski's hut found hundreds of documents written by Kaczynski but not published anywhere. An analysis produced by FBI Supervisory Special Agent James R. Fitzgerald identified numerous lexical items and phrases common to the two documents. Some were more distinctive than others, but the prosecution argued that even the more common words and phrases being used by Kaczynski became distinctive when used in combination with one another.

## **Discourse analysis**

Discourse analysis deals with analyzing written, oral, or sign language use, or any significant semiotic event. According to the method, the close analysis of a covert recording can produce useful deductions. The use of "I" instead of "we" in a recording highlights non-complicity in a conspiracy. The

utterance of "yeah" and "uh-huh" as responses indicate that the suspect understands the suggestion, while feedback markers such as "yeah" and "uh-huh" do not denote the suspect's agreement to the suggestion. Discourse analysts are not always allowed to testify but during preparation for a case they are often useful to lawyers.

## **Linguistic dialectology**

This refers to the study of dialects in a methodological manner based on anthropological information. It is becoming more important to conduct systematic studies of dialects, especially within the English language, because they are no longer as distinct as they once were due to the onslaught of mass media and population mobility. Political and social issues have also caused languages to straddle geographical borders resulting in certain language varieties spoken in multiple countries, leading to complications when determining an individual's origin by means of his/her language or dialect.

Dialectology was used during the investigations into the Yorkshire Ripper tape hoax.

## **Forensic phonetics**

The forensic phonetician is concerned with the production of accurate transcriptions of what was being said. Transcriptions can reveal information about a speaker's social and regional background. Forensic phonetics can determine similarities between the speakers of two or more separate recordings. Voice recording as a supplement to the transcription can be useful

as it allows victims and witnesses to indicate whether the voice of a suspect is that of the accused, i.e. alleged, criminal.

A man accused of manufacturing the drug ecstasy was misheard by the police transcriber as "hallucinogenic". The police transcriber heard "but if it's, as you say, it's hallucinogenic, it's in the Sigma catalogue". However, the actual utterance was "but if it's, as you say, it's German, it's in the Sigma catalogue".

Another disputed utterance was between a police officer and a suspect. One of the topics of conversation was a third man known as "Ernie". The poor signal of the recording made "Ernie" sound like "Ronnie". The surveillance tape presented acoustic problems: an intrusive electronic-sounding crackle, the sound of the car engine, the playing of the car radio, the movement of the target vehicle, and the intrusive noise all coincided with the first syllable of the disputed name.

Forensic speechreading is the complement of forensic voice identification. Transcripts of surveilled video records can sometimes allow expert speechreaders to identify speech content or style where the identity of the talker is apparent from the video record.

## **Examples**

Evidence from forensic linguistics has more power to eliminate someone as a suspect than to prove him or her guilty. Linguistic expertise has been employed in criminal cases to defend an individual suspected of a crime and during government investigations. Forensic linguists have given expert

evidence in a wide variety of cases, including abuse of process, where police statements were found to be too similar to have been independently produced by police officers; the authorship of hate mail; the authorship of letters to an Internet child pornography service; the contemporaneity of an arsonist's diary; the comparison between a set of mobile phone texts and a suspect's police interview, and the reconstruction of a mobile phone text conversation. Some well-known examples include an appeal against the conviction of Derek Bentley; the identification of Subcomandante Marcos, the Zapatistas' charismatic leader, by Max Appedole; and the identification of Theodore Kaczynski as the so-called "Unabomber" by James R. Fitzgerald.

The criminal laboratories *Bundeskriminalamt* (in Germany) and the *NederlandsForensischInstituut* (in the Netherlands) both employ forensic linguists.

Forensic linguistics contributed to the overturning of Derek Bentley's conviction for murder in 1998, although there were other non-linguistic issues. Nineteen-year-old Bentley, who was functionally illiterate, had been hanged in 1953 for his part in the murder of PC Sidney Miles; he had been convicted partly on the basis of his statement to police, allegedly transcribed verbatim from a spoken monologue. When the case was reopened, a forensic linguist found that the frequency and usage of the word "then" in police transcripts suggested the transcripts were not verbatim statements but had been partially authored by police interviewers; this and other evidence led to Bentley's posthumous pardon.

During the investigative stage in identifying Subcomandante Marcos, the Mexican government speculated that he was a dangerous guerrilla fighter. This theory gained much traction at the end of 1994, after the dissident Zapatista Comandante, Salvador Morales Garibay, gave away the identity of his former fellow Zapatistas to the Mexican government, among them Marcos' identity. They all were indicted for terrorism, arrest warrants were issued, and arrests were made in a military action. The Mexican government alleges some Zapatistas to be terrorists, among them Marcos. There was a storm of political pressures claiming for a fast military solution to the 1995 Zapatista Crisis. On 9 February 1995, in a televised special presidential broadcast, President Ernesto Zedillo announced Subcomandante Marcos to be one Rafael Sebastián Guillén Vicente, born 19 June 1957 to Spanish immigrants in Tampico, Tamaulipas, a former professor at the Universidad Autónoma Metropolitana School of Sciences and Arts for the Design.

After the government revealed Marcos' identity in January 1995, Max Appedole, old friend, classmate with the Jesuits at the Instituto Cultural Tampico, made direct intervention in the conflict. Appedole played a major role with the Mexican government to avoid a military solution to the 1995 Zapatista Crisis by demonstrating that, contrary to the accusations announced by President Ernesto Zedillo, Rafael Guillén was no terrorist. Max Appedole identified Marcos' linguistic fingerprint, based on Marcos' specific, unique way of speaking, recognized his literary style in all Marcos' manifestos that were published in the media, and linked them to literary tournaments organized by the Jesuits in which they competed in Mexico. Everyone has an idiolect, encompassing vocabulary,

grammar, and pronunciation, that differs from the way other people talk. He confirmed that he had no doubt that Marcos was his friend Rafael Guillén, a pacifist. Max Appedole closed the first successful linguistic profiling confirmation case in the history of law enforcement. Based on these achievements, a new science was developed, giving way to what is now called forensic linguistics. This motivated a new division of forensic linguistics called "criminal profiling in law enforcement".

Forensic linguistic evidence also played a role in the investigation of the 2005 disappearance of Julie Turner, a 40-year-old woman living in Yorkshire. After she was reported missing, her partner received several text messages from Julie's mobile phone, such as "Stopping at jills, back later need to sort my head out", and "Tell kids not to worry. sorting my life out. be in touch to get some things" [*sic*]. Investigators found that letters written by Turner's friend Howard Simmerson shared linguistic similarities with the text messages, suggesting that Simmerson had been aware of the contents of the messages. Simmerson was eventually found guilty of Turner's murder.

19-year-old Jenny Nicholl disappeared on 30 June 2005. Her body was never found, giving police and forensic scientists little information to go on about what might have happened to Jenny. After looking through her phone for clues, forensic linguists came to the conclusion that the texts sent from her phone around the time that she disappeared seemed very different from her usual texting style, and soon started looking to her ex-boyfriend, David Hodgson, for clues of what happened to her, including looking through his phone and studying his texting style. The forensic linguists found a

number of stylistic similarities between David's texting style and the messages sent from Jenny's phone around the time she went missing. Using the timeframe of when she went missing, combined with the differences in texting styles and other forensic details, Jenny's murderer, David Hodgson, was convicted. The analysis of the text messages and their submission in court helped to pave the way for forensic linguistics to be acknowledged as a science in UK law, rather than opinion. To this day, her body has not been found, but justice was still served for her and her family because of forensic linguistics.

Forensic linguist John Olsson gave evidence in a murder trial on the meaning of "jooking" in connection with a stabbing.

During the appeal against the conviction of the Bridgewater Four, the forensic linguist examined the written confession of Patrick Molloy, one of the defendants – a confession which he had retracted immediately – and a written record of an interview which the police claimed took place immediately before the confession was dictated. Molloy denied that the interview had ever taken place, and the analysis indicated that the answers in the interview were not consistent with the questions being asked. The linguist came to the conclusion that the interview had been fabricated by police. The conviction against the Bridgewater Four was quashed before the linguist in the case, Malcolm Coulthard, could produce his evidence.

In an Australian case reported by Eagleson, a "farewell letter" had apparently been written by a woman prior to her disappearance. The letter was compared with a sample of her



previous writing and that of her husband. Eagleson came to the conclusion that the letter had been written by the husband of the missing woman, who subsequently confessed to having written it and to having killed his wife. The features analyzed included sentence breaks, marked themes, and deletion of prepositions.

In 2009, there was a house fire where a father was able to save his children, but his wife died in the house. The police thought that the fire was actually not an accident, but instead a cover-up of the father murdering the mother. The forensic linguists were able to obtain the phones of both the father and mother, and realized that there were texts still being sent from the mother's phone the whole day – long after the police thought she had died. Using information from the two phones, the linguists were able to study the texting styles of both parents to see if they could obtain any more information about what happened that day. It turned out that the texts sent from the wife's phone were actually the husband pretending to be the wife so that no one would know she was murdered, and everyone would believe that she perished in the house fire. The forensic linguists were able to figure this out by studying the husband's texting style, spelling errors, and more, and were able to come to the conclusion that the texts sent after the wife was thought to be deceased were actually the husband texting from her phone pretending to be her. Without this knowledge, it would have been much more difficult to convict the husband of murder and get justice for the family.

# **Additional concepts**

## **Linguistic fingerprinting**

A linguistic fingerprint is a concept put forward by some scholars that each human being uses language differently, and that this difference between people involves a collection of markers which stamps a speaker/writer as unique, similar to a fingerprint. Under this view, it is assumed that every individual uses languages differently and this difference can be observed as a fingerprint. It is formed as a result of merged language style. A person's linguistic fingerprint can be reconstructed from the individual's daily interactions and relate to a variety of self-reported personality characteristics, situational variables, and physiological markers (e.g. blood pressure, cortisol, testosterone). In the process of an investigation, the emphasis should be on the relative rather than absolute difference between the authors and how investigators can classify their texts. John Olsson, however, argues that although the concept of linguistic fingerprinting is attractive to law enforcement agencies, there is so far little hard evidence to support the notion.

## **Variation**

Intra-author variations are the ways in which one author's texts differ from each other. Inter-author variations are the ways in which different authors' writing varies. Two texts by one author do not necessarily vary less than texts by two different authors.

- **Genre:** When texts are being measured in different genres, considerable variation is observed even though they are by the same author.
- **Text Type:** Personal letters contain more inter-relationship bonding strategies than academic articles or term papers.
- **Fiction vs. Non-Fiction:** Some fiction writers are journalists. Due to the different demands of each medium, they can be completely different from one another and this results in intra-author variation.
- **Private vs. Public:** A politician writing a political speech, which is a public text, will differ greatly from a private text to a friend or family member.
- **Time lapse:** The greater the time lapse between two works, the greater the likely variation. Language changes more than we realize in a relatively short span of time, influencing our susceptibility to language changes around us.
- **Disguise:** A writer can publish pseudonymously or anonymously, disguising output to prevent recognition.

## **Forensic transcription**

The two main types of transcriptions are written documents and video and audio records. Accurate, reliable text transcription is important because the text is the data which becomes the available evidence. If a transcription is wrong, the evidence is altered. If there is failure to transcribe the full text, evidence is once again altered unwittingly. There must be emphasis on the text being the evidence. A transcription of an

audio file should never be assumed to be completely accurate. Each type of transcription contains its own problems. A handwritten document might contain unusual spellings which may result in ambiguous meanings, illegible handwriting, and illustrations that are difficult to comprehend. A scanned document is tricky, as scanning may alter the original document. Audio and video documents can include repetitions, hesitation, nonsensical talk, jargon which can be hard to understand, and speakers mumbling incoherently and inaudibly. Non-linguistic sounds such as crying and laughing may also be included in the audio and video text which cannot be transcribed easily. Because of this, civil libertarians have argued that interrogations in major criminal cases should be recorded and the recordings kept, as well as transcribed.

## **Forensic speechreading**

**Forensic speechreading** (or **forensic lipreading**) is the use of speechreading for information or evidential purposes. Forensic speechreading can be considered a branch of forensic linguistics. In contrast to speaker recognition which is often the focus of voice analysis from an audio record, forensic speechreading usually aims to establish the *content* of speech, since the *identity* of the talker is usually apparent. Often, it involves the production of a transcript of lipread video-records of talk that lack a usable audiotrack, for example CCTV material. Occasionally, 'live' lipreading is involved, for example in the Casey Anthony case. Forensic speechreaders are usually deaf or from deaf families (CODA), and use speechreading in their daily lives to a greater extent than people with normal hearing outwith a deaf community. Some speechreading tests

suggest deaf people can be better lipreaders than most hearing people.

## **Speechreading expertise**

No tests of speechreading have yet been developed in a forensic context: that is, to benchmark individual skills in speechreading from a video record, including production of a reliable transcript. For many years, UK agencies made extensive use of one particular speechreader, whose reports are now not to be used for evidential purposes. Several speechreaders and deaf professionals currently offer these services. In the UK, these include Terry Ruane, Laraine Callow MBE, Tina Lannin among others. In the US, Consuelo Gonzalez is a leading lipreading expert. Expert speechreaders may be able to advise on various issues, including whether a videorecord is or is not speechreadable, and the accent and language used by a talker. Commissioning agents need to be aware of issues inherent in the unreliability of speechreading, and be prepared to treat such advice with caution.

## **The law**

In the UK, a landmark case and appeal (*R. v Luttrell et al.*, 2004) established the admissibility of lipreading evidence. However, the appeal court also required that the judge should issue a special warning as to its risks and limitations.

While lipread speech can carry useful speech information, it is inherently less accurate than (clearly) heard speech because

many distinctive features of speech are produced by actions of the tongue within the oral cavity and are not visible. This is a limitation imposed by speech itself, not the expertise of the speechreader. It is the main reason why the accuracy of a speechreader working on a purely visual record cannot be considered wholly reliable, however skilled they may be and irrespective of hearing status. The type of evidence and the utility of such evidence varies from case to case.

In the US, there is debate concerning the admissibility of speechreading evidence and its status, especially in relation to variations in state and federal evidential procedures, and with respect to the privacy implications of the Fourth Amendment to the US constitution.

## **Three UK cases involving speechreading evidence**

- **R v Luttrell.** Nine defendants, including Luttrell, were found guilty on a charge of conspiracy to commit armed robbery and dispose of stolen goods. Before his arrest, Luttrell, who pleaded not guilty, was put under surveillance by the police, and a video record was obtained of a conversation. The expert lipreader, acting for the prosecution, produced a transcript of the conversation, indicating that he spoke the name of a co-conspirator. Although the defence claimed this was unreliable evidence, the judge ruled it admissible. On appeal, Luttrell's conviction was upheld, but the court of appeal made

clear that juries should be warned of the unreliability of such evidence.

- **Nat Fraser.** The expert lipreader's transcript of a surveilled prison conversation between Nat Fraser and his friend, Glenn Lucas, was used by the prosecution to help secure a conviction for Fraser's murder of his wife, Arlene Fraser (Murder of Arlene Fraser). That verdict was overturned in part on the basis of unreliability of that lipreader's evidence. Following successful appeal against his original conviction, on re-trial in 2012 Fraser was again convicted of his wife's murder (*Fraser v HM Advocate*).
- **R v John Terry.** John Terry was alleged to have made a crude racist remark against Anton Ferdinand in the course of a televised Premier League football match between Chelsea FC and Queen's Park Rangers FC. Terry's remarks were seen in ten seconds of widely broadcast and re-broadcast close-up, and he was prosecuted under public order crime provisions, which include offences which are racially aggravated. Terry's defence argued that the remark was made in response to an earlier accusation from Ferdinand; Terry was just "sarcastically repeating a racist slur" he thought Anton Ferdinand had wrongly accused him of making, to refute the perceived allegation and to establish what had been said. In a joint report, two expert speechheaders corroborated the content of Terry's broadcast remarks, but could not infer why they were said, since there was no video record of the preceding context. The magistrate ruled that Terry could not be found guilty because of

this uncertainty. The Football Association, whose Independent Regulatory Commission examined the case with a view to disciplinary action, came to a different conclusion. The Commission dismissed the lip readers' evidence, claiming "it is to be observed that the Chief Magistrate did not derive any real assistance from the evidence of the two lip readers who were called to give evidence", although the Chief Magistrate had himself spoken of the importance of the lip reading evidence: "Other words appear to be spoken. Both parties have agreed that expert evidence from lip readers is necessary to say what those words are. It is axiomatic that expert evidence is not called unless a particular expertise is needed to give an opinion to the court which the court cannot readily form itself." The Commission based its decision (which required a lower standard of proof) on a transcription of the video evidence that conflicted with the evidence of the lip readers, as well as on disputed circumstantial evidence and references to the disputed post-incident behaviour of Anton Ferdinand, John Terry, Ashley Cole and David Barnard.

## **Native-language identification**

**Native-language identification (NLI)** is the task of determining an author's native language (L1) based only on their writings in a second language (L2). NLI works through identifying language-usage patterns that are common to specific L1 groups and then applying this knowledge to predict



the native language of previously unseen texts. This is motivated in part by applications in second-language acquisition, language teaching and forensic linguistics, amongst others.

## **Overview**

NLI works under the assumption that an author's L1 will dispose them towards particular language production patterns in their L2, as influenced by their native language. This relates to cross-linguistic influence (CLI), a key topic in the field of second-language acquisition (SLA) that analyzes transfer effects from the L1 on later learned languages.

Using large-scale English data, NLI methods achieve over 80% accuracy in predicting the native language of texts written by authors from 11 different L1 backgrounds. This can be compared to a baseline of 9% for choosing randomly.

## **Applications**

### **Pedagogy and language transfer**

This identification of L1-specific features has been used to study language transfer effects in second-language acquisition. This is useful for developing pedagogical material, teaching methods, L1-specific instructions and generating learner feedback that is tailored to their native language.

## **Forensic linguistics**

NLI methods can also be applied in forensic linguistics as a method of performing authorship profiling in order to infer the attributes of an author, including their linguistic background. This is particularly useful in situations where a text, e.g. an anonymous letter, is the key piece of evidence in an investigation and clues about the native language of a writer can help investigators in identifying the source. This has already attracted interest and funding from intelligence agencies.

## **Methodology**

Natural language processing methods are used to extract and identify language usage patterns common to speakers of an L1-group. This is done using language learner data, usually from a learner corpus. Next, machine learning is applied to train classifiers, like support vector machines, for predicting the L1 of unseen texts. A range of ensemble based systems have also been applied to the task and shown to improve performance over single classifier systems.

Various linguistic feature types have been applied for this task. These include syntactic features such as constituent parses, grammatical dependencies and part-of-speech tags. Surface level lexical features such as character, word and lemma n-grams have also been found to be quite useful for this task. However, it seems that character n-grams are the single best feature for the task.

## **2013 shared task**

The Building Educational Applications (BEA) workshop at NAACL 2013 hosted the inaugural NLI shared task. The competition resulted in 29 entries from teams across the globe, 24 of which also published a paper describing their systems and approaches.