Sociolinguistics, Computational Linguistics

DARSHIL MAKWANA
Research Scholar,
Sentral University of Gujarat (CUG), Gandhinagar

1. Introduction

Sociolinguistics refers to the way language is used in society. A dialect is a variety of language that is systematically different from other varieties of the same language. The dialects of a single language are mutually intelligible, but when the speakers can no longer understand each other, the dialects become languages. Geographical regions are also considered when dialects become languages. The discussion or the understanding of the relationship between language and society, or of the various functions of language in society, should begin with some attempt to define each of the terms. Let us say that a society is any group of people who are drawn together for a certain purpose or purposes. By such a definition 'society' becomes a very comprehensive concept. The comprehensive definition of language: a language is what the members of a particular society speak. However, as we will see, speech in almost any society can take many very different forms, and just what forms we should choose to discuss when we attempt to describe the language of a society may prove to be a contentious matter. Besides dialects, speakers may use different styles or registers (such as contractions) depending on the context. Slang may also be used in speech, but is not often used in formal situations or writing. Jargon refers to the unique vocabulary pertaining to a certain area, such as computers or medicine. Words or expressions referring to certain acts that are forbidden or frowned upon are considered taboo. These taboo words produce euphemisms, words or phrases that replace the expressions that are being avoided. Sometimes too a society may be plurilingual; that is, many speakers may use more than one language, however we define language. We should also note that our definitions of language and society are not independent: the definition of language includes in it a reference to society. I will return to this matter from time to time. When two or more people communicate with each other in speech, we can call the system of communication that they employ a code. In most cases that code will be something we may also want to call a language. In practice, linguists do not find it at all easy to write grammars because the knowledge that people have of the languages, they speak is extremely hard to describe. It is certainly something different from, and is much more considerable than, the kinds of knowledge we see described in most of the grammars we find on library shelves, no matter how good those grammars may be. The use of words may indicate a society's attitude toward sex, bodily functions or religious beliefs, and they may also reflect racism or sexism in a society. Language itself is not racist or sexist, but the society may be. Such insulting words may reinforce biased views, and changes in society may be reflected in the changes in language.

2. Objectives

1. To get an insight into the significant relationship between language and society.
2. To understand and define sociolinguistics as an independent discipline.
3. To trace the growth of sociolinguistics for the recent research work.
4. To appreciate and get familiar with the great diversity of phenomena that sociolinguists investigate, for instance, language variation, language change, language maintenance, language planning, etc.

3. Extension of the meaning

3.1 Language and Society

In the history of linguistics, it is rare to find investigations of any language which are entirely cut off from concurrent investigations of the history of that language, or of its regional and/or social

distributions, or of its relationship to objects, ideas, events, and actual speakers and listeners in the 'real' world. That is one of the reasons why a number of linguists have found Chomsky's asocial view of linguistic theorizing to be a rather sterile type of activity, since it explicitly rejects any concern for the relationship between a language and those who use it. Hudson (1996) calls linguistic items, 'such entities as sounds, words, grammatical structures, and so on are essential in any language. It is these items, their status, and their arrangements that language theorists such as Chomsky concern themselves with. On the other hand, social theorists, particularly sociologists, attempt to understand how societies are structured and how people manage to live together. To do so, they use such concepts as identity, power, class, status, solidarity, accommodation, face, gender, politeness, etc. A major concern of is to examine possible relationships between linguistic items on the one hand and concepts such as 'power', 'solidarity', 'etc. on the other. Linguistic items are difficult to define. Try, for example, to define exactly what linguistic items such as sounds, syllables, words, and sentences are. Then try to define precisely what you understand by such concepts as social class, solidarity, identity, face, and politeness. Finally, try to relate the two sets of definitions within some kind of theory so as to draw conclusions about how items in these two very different classes relate to each other. Do all this while keeping in mind that languages and societies are constantly changing.

There are several possible relationships between language and society. One is that social structure may either influence or determine linguistic structure and/or behavior. The varieties of language that speakers use reflect such matters as their regional, social, or ethnic origin and possibly even their gender; and other studies which show that particular ways of speaking, choices of words, and even rules for conversing are in fact highly determined by certain social requirements. A possible relationship is directly opposed to the first: linguistic structure and/or behavior may either influence or determine social structure. The other possible relationship is that the influence is bi-directional: language and society may influence each other. Next is to assume that there is no relationship at all between linguistic structure and social structure and that each is independent of the other. Actually, this variant view appears to be the one that Chomsky himself holds: he prefers to develop an asocial linguistics as a preliminary to any other kind of linguistics, such an asocial approach being, in his view, logically prior.

3.2 Languages, Dialects, and Varieties

Hudson (1996, p. 22) defines a variety of language as a set of linguistic items with similar distribution, a definition that allows us to say that all of the following are varieties: Canadian English, London English, the English of football commentaries, and so on. According to Hudson, this definition also allows us to treat all the languages of some multilingual speaker, or community, as a single variety, since all the linguistic items concerned have a similar social distribution. 'A variety can therefore be something greater than a single language as well as something less, less even than something traditionally referred to as a dialect. Ferguson (1972, p. 30) offers another definition of variety: anybody of human speech patterns which is sufficiently homogeneous to be analyzed by available techniques of synchronic description and which has a sufficiently large repertory of elements and their arrangements or processes with broad enough semantic scope to function in all formal contexts of communication. Hudson and Ferguson agree in defining variety in terms of a specific set of _linguistic items or human speech patterns (presumably, sounds, words, grammatical features, etc.) which we can uniquely associate with some external factor (presumably, a geographical area or a social group). Consequently, if we can identify such a unique set of items or patterns for each group in question, it might be possible to say there are such varieties as Standard English, Cockney, lower-class New York City speech, Oxford English, legalese, cocktail party talk, and so on. One important task, then, in sociolinguistics is to determine if such unique sets of items or patterns do exist. As we proceed, we will encounter certain difficulties, but it is unlikely that we will easily abandon the concept of variety, no matter how serious these difficulties prove to be.

3.3 Language and Dialect

For many people there can be no confusion at all about what language they speak. Most speakers can give a name to whatever it is they speak. On occasion, some of these names may appear to be strange to those who take a scientific interest in languages, but we should remember that human naming practices often have a large unscientific component to them. Census-takers in India find themselves confronted with a wide array of language names when they ask people what language or languages they speak. Names are not only ascribed by region, which is what we might expect, but sometimes also by caste, religion, village, and so on. Moreover, they can change from census to census as the political and social climate of the country changes. For example, the same word or nonverbal communication in India differs from place to place. The word *dada* in Gujarat means grandfather but the same is not so in East India. Nonverbal communication of nodding the head vertically in West India means No, but the same is used for agreement in North- East India. While people do usually know what language they speak, they may not always lay claim to be fully qualified speakers of that language. They may experience difficulty in deciding whether what they speak should be called a language proper or merely a dialect of some language. Such indecision is not surprising: exactly how do you decide what is a language and what is a dialect of a language? What criteria can you possibly use to determine that, whereas variety X is a language, variety Y is only a dialect of a language? What are the essential differences between a language and a dialect? Haugen (1966a) has pointed out that language and dialect are ambiguous terms. Ordinary people use these terms quite freely in speech; for them a dialect is almost certainly no more than a local non-prestigious (therefore powerless) variety of a real language. In contrast, scholars often experience considerable difficulty in deciding whether one term should be used rather than the other in certain situations. As Haugen says, the terms represent a simple dichotomy in a situation that is almost infinitely complex. He points out that the confusion goes back to the Ancient Greeks. The Greek language that we associate with Ancient Greece was actually a group of distinct local varieties (Ionic, Doric, and Attic) descended by divergence from a common spoken source with each variety having its own literary traditions and uses, e.g., Ionic for history, Doric for choral and lyric works, and Attic for tragedy. Later, Athenian Greek, the koiné – or common language – became the norm for the spoken language as the various spoken varieties converged on the dialect of the major cultural and administrative center.

Dialect is used both for local varieties of English, e.g., Various Gujarati dialects, and for various types of informal, lower-class, or rural speech. In general usage it therefore remains quite undefined whether such dialects are part of the language‖ or not. In fact, the dialect is often thought of as standing outside the language… As a social norm, then, a dialect is a language that is excluded from polite society (pp. 924–5). It is often equivalent to nonstandard or even substandard, when such terms are applied to language, and can connote various degrees of inferiority, with that connotation of inferiority carried over to those who speak a dialect. We can observe too that questions such as Which language do you speak? or Which dialect do you speak? may be answered quite differently by people who appear to speak in an identical manner. Every dialect has its own unique way to be understood and spoken. The difference can be seen as a person moves from his/her native to an urban area.

The various relationships among languages and dialects can be used to show how the concepts of power and solidarity help us understand what is happening. Power requires some kind of asymmetrical relationship between entities: one has more of something that is important, e.g., status, money, influence, etc., than the other or others. A language has more power than any of its dialects. It is the powerful dialect but it has become so because of non-linguistic factors. Standard English and Parisian French are good examples. Solidarity, on the other hand, is a feeling of equality that people have with one another. They have a common interest around which they will bond. A feeling of solidarity can lead people to preserve a local dialect or an endangered language to resist power, or to insist on independence.

We undoubtedly agree that whatever is written in English and that English is a language, but we may be less certain that various other things we see written or hear spoken in what is called English should properly be regarded as English rather than as dialects or varieties of English, perhaps variously described as Indian English, Australian English, New York English, West Country English, African American Vernacular English, nonstandard English, BBC English, and so on. A language then would be some unitary system of linguistic communication which subsumes a number of mutually intelligible varieties. It would therefore be bigger than a single dialect or a single variety. However, that cannot always be the case, for some such systems used by very small numbers of speakers may have very little internal variation. Yet each must be a language, for it is quite unlike any other existing system. Actually, neither the requirement that there be internal variation nor the numbers game, i.e., that a language must somehow be bigger than a dialect, offers much help. Many languages have only a handful of speakers; several have actually been known to have had only a single remaining speaker at a particular point in time and the language has died with that speaker.

3.4 Computational Lingusitics

Computational linguistics is the scientific and engineering discipline concerned with understanding written and spoken language from a computational perspective, and building artifacts that usefully process and produce language, either in bulk or in a dialogue setting. To the extent that language is a mirror of mind, a computational understanding of language also provides insight into thinking and intelligence. And since language is our most natural and most versatile means of communication, linguistically competent computers would greatly facilitate the interaction. The theoretical goals of computational linguistics include the formulation of grammatical and semantic frameworks for characterizing languages in ways enabling computationally tractable implementations of syntactic and semantic analysis.

The methods employed in theoretical and practical research in computational linguistics have often drawn upon theories and findings in theoretical linguistics, philosophical logic, cognitive science (especially psycholinguistics), and of course computer science. However, early work from the mid-1950s to around 1970 tended to be rather theory-neutral, the primary concern being the development of practical techniques for such applications as MT (Machine Translation) and simple QA (Question Answering). In MT, central issues were lexical structure and content, the characterization of sublanguages‖ for particular domains (for example, weather reports), and the transduction from one language to another (for example, using rather ad hoc graph transformation grammars or transfer grammars). In QA, the concern was with characterizing the question patterns encountered in a specific domain, and the relationship of these question patterns to the forms in which answers might store, for instance in a relational database.

3.5 Syntax

Before considering how grammatical structure can be represented, analyzed and used, we should ask what basis we might have for considering a particular grammar correct, or a particular sentence grammatical, in the first place. Of course, these are primarily questions for linguistics proper, but the answers we give certainly have consequences for computational linguistics.

Traditionally, formal grammars have been designed to capture linguist's intuitions about well-formedness as concisely as possible, in a way that also allows generalizations about a particular language (e.g., subject-auxiliary inversion in English questions) and across languages (e.g., a consistent ordering of nominal subject, verb, and nominal object for declarative, pragmatically neutral main clauses). Concerning linguists' specific well-formedness judgments, it is worth noting that these are largely in agreement not only with each other, but also with judgments of non-linguists at least for clearly grammatical and clearly ungrammatical‖ sentences (Pinker 2007). Also, the discovery that conventional phrase structure supports elegant compositional theories of meaning lends credence to the traditional theoretical methodology.

However, traditional formal grammars have generally not covered any one language comprehensively, and have drawn sharp boundaries between well-formedness and ill-formedness, when in fact people's (including linguists') grammaticality judgments for many sentences are uncertain or equivocal. Moreover, when we seek to process sentences in the wild‖, we would like to accommodate regional, genre-specific, and register-dependent variations in language, dialects, and erroneous and sloppy language (e.g., misspellings, unpunctuated run-on sentences, hesitations and repairs in speech, faulty constituent orderings produced by non-native speakers, and fossilized errors by native speakers, such as for you and I possibly a product of school teachers inveighing against you and me‖ in subject position). Consequently, linguists' idealized grammars need to be made variation-tolerant in most practical applications. The way this need has typically been met is by admitting a far greater number of phrase structure rules than linguistic parsimony would sanction say, 10,000 or more rules instead of a few hundred. These rules are not directly supplied by linguists (computational or otherwise), but rather can be read off‖ corpora of written or spoken language that have been decorated by trained annotators (such as linguistics graduate students) with their basic phrasal tree structure. Unsupervised grammar acquisition (often starting with POS-tagged training corpora) is another avenue, but results are apt to be less satisfactory. In conjunction with statistical training and parsing techniques, this loosening of grammar leads to a rather different conception of what comprises a grammatically flawed sentence: It is not necessarily one rejected by the grammar, but one whose analysis requires some rarely used rules. Yule (2010) defines syntax as: Syntax concerns the way that words are arranged into larger units. That is, words are the basic units the building blocks of syntactic analysis. The largest unit that syntactic analysis usually considers is the sentence. For this reason, syntax is often equated with the study of sentence structure.

A finite number of rules facilitate an infinite number of sentences that can be simultaneously understood by both the speaker and the listener. In order for this to work with any degree of success, the rules have to be precise and have to be consistently adhered to. The grammar of a language has several components.
These can be described as follows:
a) The phonetics that governs the structure of sounds;
b) The morphology that governs the structure of words;
c) The syntax, which governs the structure of sentences
d) The semantics that governs the meanings of words and sentences.

3.6 Tree Diagram
One of the most common ways to create a visual representation of syntactic structure is through tree diagrams. Symbols (Art=article, N = noun, NP = noun phrase) are used to label the parts of the tree to capture the hierarchical organization of those parts in the underlying structure of phrases and sentences. There are symbols that are used as abbreviations for syntactic categories. Examples are- S (= sentence), ―NP (= noun phrase), ―N (= noun), ―Art (= article), ―V (= verb) and ―VP (= verb phrase), ―PP (= prepositional phrase). There are three more symbols that are commonly used in syntactic description. Yule (2010) gives detail of these symbols as:

(i) The first is in the form of an arrow →. It can be interpreted as consists of‖. For example, NP → Art N

(ii) The second symbol is a pair of round brackets ( ). Whatever occurs inside these round brackets will be treated as an optional constituent. For example, NP → Art (Adj) N
(iii) The third symbol is in the form of curly brackets {}. These indicates that only one of the elements enclosed within the curly brackets must be selected. We use these types of brackets when we want to indicate that there is a choice from two or more constituents.

International Journal of Research in all Subjects in Multi Languages
[Author: Darshil Makwana] [Sub.: English] I.F.6.133

Vol. 11, Issue: 6, June: 2023
(IJRSML) ISSN: 2321 - 2853

List of symbols used in tree diagram is given below.
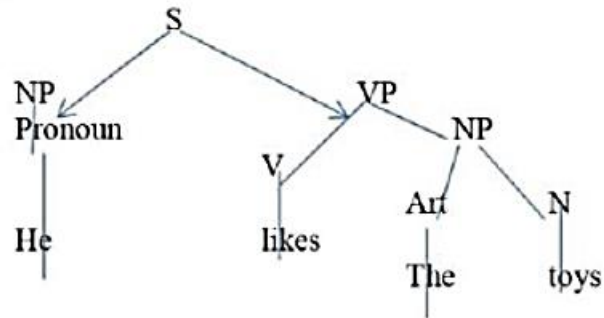= Sentence, NP = Noun phrase, PN = Proper noun
N = Noun, VP = Verb phrase, Adv = Adverb
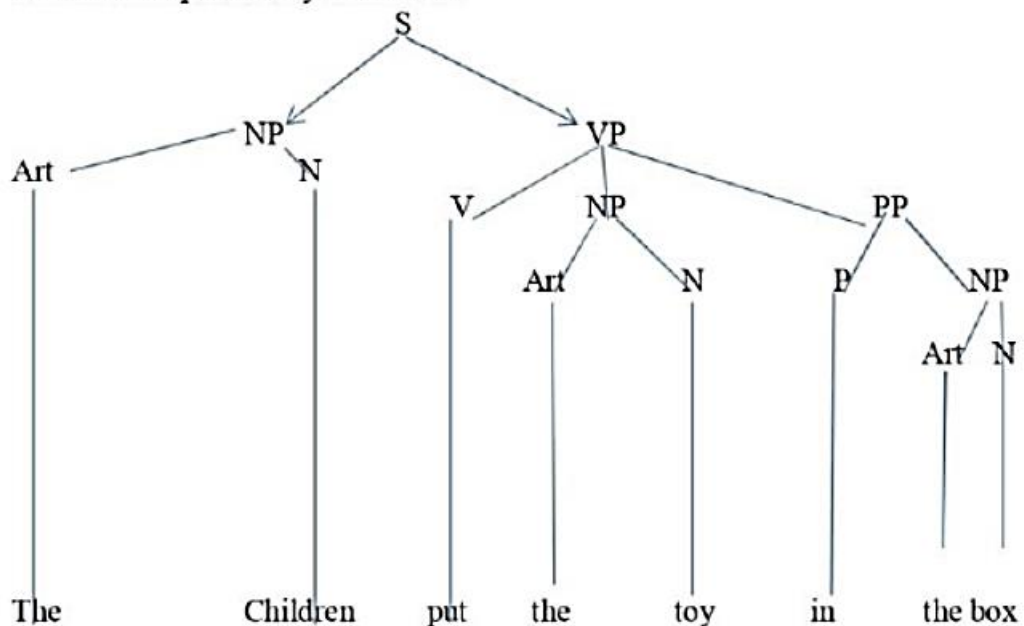V = Verb, Adj = Adjective, Prep = Preposition
Art = Article, Pro = Pronoun, PP = Prepositional phrase

Examples
1. He likes the toys.

2. The children put the toy in the box.

References
1. Bailey, R. W. and M. Gorlach (eds.) English as a World Language. Ann Arbor: University of Michigan Press. 1982
2. Bell, R. T. Sociolinguistics: Goals, Approaches and Problems. London: Batsford. 1976
3. Baker, L. English Syntax. London: The MIT Press. 1998.
4. Trask, R. Key Concepts in Language and Linguistics. New York: The Routledge Publishers. 1999