DISAMBIGUATION OF AMBIGUITIES IN ENGLISH SENTENCES

Maurice Udom, Ph.D
Department of General Studies
Akwa Ibom State Polytechnic, Ikot Osurua.

Uchenna Enyia
General Studies Unit
Topfaith University, Mkpatak, Akwa Ibom State

ABSTRACT

The intricacies of language are a captivating and complex phenomenon. It occurs when a word, phrase, or sentence can be interpreted in multiple ways, leading to confusion or difficulty grasping the intended meaning. This complexity arises from various factors, including word placement, polysemy (words with multiple meanings), and homonymy (words that sound alike but have different meanings). This research aims to comprehensively analyze and explore sentences with various meanings through a knowledge-based approach, grammatical structure analysis, and diagrams to clarify their intended interpretations and enhance their comprehensibility. It seeks to contribute significantly to our understanding of ambiguous language structures and their impact on communication. The study reveals ambiguity in every language, single words, and longer sentences. Sentences lacking structural signals and those with imprecise constituents result in syntactic or structural ambiguity. The study recommends using clear

sentence structures and incorporating appropriate signals in sentences.

Keywords: *ambiguity, polysemy, homonymy, lexical ambiguity, syntactic ambiguity.*

INTRODUCTION

Ambiguities in English sentences present a significant challenge in natural language processing (NLP) due to their potential to cause confusion and misunderstanding. The disambiguation of such ambiguities is critical for improving the accuracy and efficiency of language processing systems across various applications. Researchers have recently dedicated efforts to developing robust disambiguation models that can effectively tackle this linguistic challenge.

One prominent model in this field is the probabilistic approach, which leverages statistical information to identify and resolve ambiguities in text. Researchers such as Manning and Schutze (1999) have explored the use of probabilistic models, such as Hidden Markov Models and Conditional Random Fields, to disambiguate word senses in context. Considering the contextual information surrounding ambiguous words, these models have shown promising results in disambiguating ambiguities in English sentences.

Furthermore, the work of scholars like Jurafsky and Martin (2009) has delved into the use of machine learning techniques, such as Support Vector Machines and neural networks, for ambiguity resolution.

These models analyze large text corpora to learn patterns and associations that can aid in disambiguating language constructs.

By training on vast amounts of annotated data, machine learning models have demonstrated the ability to effectively disambiguate various ambiguities, from word senses to syntactic ambiguities.

Another notable line of research in disambiguation is using lexical resources and semantic knowledge bases. Writers such as Fellbaum (1998) have advocated utilizing resources like WordNet, which provides structured lexical information and semantic relations between words. Researchers have enhanced disambiguation accuracy by incorporating such resources into disambiguation models,

particularly in cases where context alone may not suffice to resolve ambiguities.

Mariano Sigman and Guillermo A. Cechi (2002) opine that WordNet helps understand the lexicon's global organization and that semantic links follow power-law, scale-variant behaviors typical of self-organizing networks. They further state that polysemy is one of the links in the semantic network.

WordNet is becoming popular as a resource to be used in the knowledge approach to disambiguate the meaning of polysemy words. WordNet is a lexical database developed at Princeton University for the English language. It organizes nouns, verbs, adjectives, and adverbs into the groups of synonyms and describes the relationships between these synonym groups, forming a semantic network among the words.

In 1986, Lesk Michael developed an algorithm to identify polysemy words' senses. He used the overlap of word definitions from the Oxford Advanced Learners Dictionary of Current English to disambiguate the word sense.

Despite the divergent models and the progress in disambiguation techniques, challenges persist in achieving optimal accuracy and robustness in ambiguity resolution. One of the primary problems disambiguation models face is the selection of appropriate features and representations to capture the nuances of context and linguistic structure. Different types of ambiguities, such as lexical and syntactic ambiguities, require tailored approaches for effective disambiguation, posing a challenge for developing general-purpose disambiguation models.

Therefore, this article aims to address this problem by investigating the lexical resources and semantic knowledge approach in the disambiguation of ambiguities in English sentences. Among the different Word Sense Disambiguation methods (WSD) methods, this research focuses on the knowledge-based approach and grammatical structure analysis, employing diagrams to elucidate the disambiguation process.

Wales (1989) says that "ambiguity is double (or multiple) meaning: an ambiguous expression has more than one interpretation" (p.19). It results when one form has two or more meanings.

This can be represented diagrammatically:
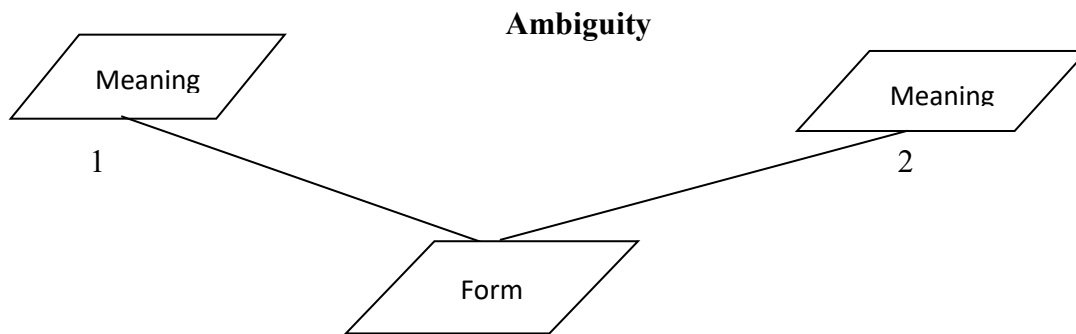
Ambiguity

The sentence, **He hit the boy with a stick**, is ambiguous because the form has two meanings:
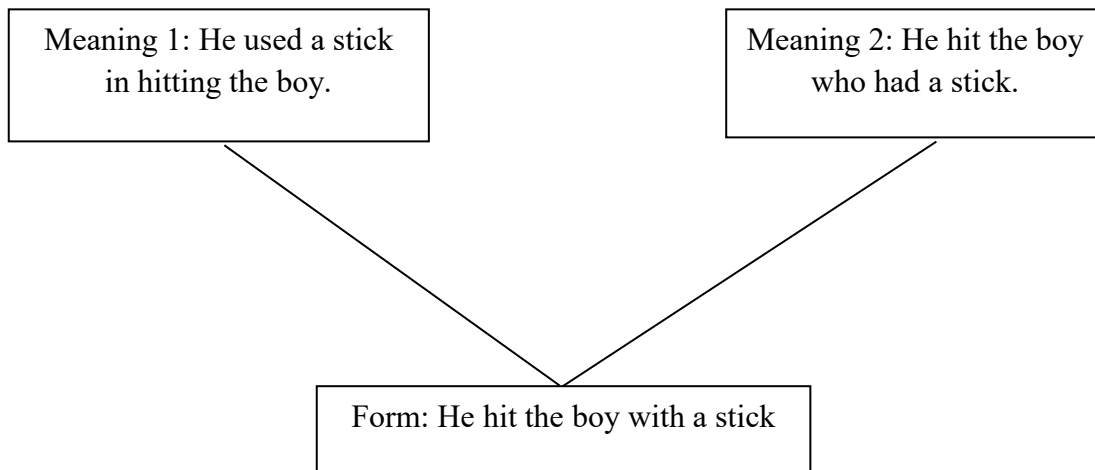
Linguists see ambiguity as a linguistic universal common to all languages and one of the inevitable consequences of the arbitrariness of language. That is the lack of one-to-one correspondence between signs and meanings.

They often distinguish between phrases or sentences' grammatical ambiguity and words' lexical ambiguity. Grammatically ambiguous units admit that the possibility of more than one structural interpretation, e.g., 'free women' as an imperative in a slogan, can also be seen as an adjective plus noun. In rhetoric, this would have been termed 'amphibologia.' If ambiguity occurs in discourse, it may not be tolerated when it hinders interpretation seriously. One of the rules arising

from the general cooperative principle of conversation (and indeed of writing) is clarity of meaning. Newspaper headlines are, sometimes, ambiguous in this 'derogatory sense,' e.g., "**British teachers amongst poorest in Europe**." Ambiguity, in this sense, is regarded as a fault of style akin to vagueness and obscurity.

Ambiguity is exploited in literature, especially in poetry, and has been seen as one of its (poetry) memorable or defining features. In poetry, the reader is not expected to be deceived or misled but to hold the different interpretations in mind and to give them equal, meaningful, severe value. Poetic ambiguity includes puns, double syntax, and any expression allowing alternative reactions or associations. In this sense, it means a multiplicity of meanings.

People recognize and tolerate ambiguity in poetry because they are attuned to accepting deviant usages and interpretations; for example, the line '**I made my song a coat**' begins Yeasts' poem "*A Coat*." This has a homonymy of two grammatical constructions:

- Subject + Verbal + Indirect Object + Direct Object

- Subject + Verbal + Direct Object + Object Complement

The first reading is equivalent to "**I made a coat for my song**," whereas the second is equivalent to "**I made my song into a coat**." These interpretations have an element of absurdity, and for this reason, both must be reckoned with in interpreting the poem.

Ambiguity – deliberate and undeliberate – is one of the main characteristics of human language. Karpf (1986) has it that "undeliberate ambiguity due to form clashes, rule conspiracy or polysemous lexical items generally escapes detection because of a safeguarding linguistic and extra-linguistic context" (p.157).

However, deliberate ambiguity in the form of puns, verbal jokes, or riddles is aimed at being detected, analyzed, and appreciated by a competent speaker of a language. The ability to perform these perceptive tasks does not solely depend on the linguistic skill of the speaker but also mainly on his metalinguistic knowledge.

Depending on the reason for the lack of specification, there are two

main types of ambiguity that linguists generally accept. These are lexical ambiguity and structural or syntactic ambiguity.

LEXICAL AMBIGUITY

There are two kinds of lexical ambiguity: polysemy and homonymy. The criteria linguists and lexicographers use in drawing the distinction between polysemy and homonymy are the etymological information of words and unrelatedness versus relatedness of meaning. Polysemy is generally detected faster than homonymy because of the shared semantic pathways and identical morphological structure.

POLYSEMY

Linguistic geographers, according to Ullman (1962), "often talk of semantic overload, hypertrophy or plethora of meaning as causes of ambiguity and confusion in language" (p.168). These terms suggest that the more senses a word has, the more ambiguous it becomes. Polysemy is a fundamental feature of human speech that can arise in many ways: shifts in application, specialization in a social milieu, foreign influence, etc. Polysemy is not a defect of language but an essential condition of its efficiency.

If it were not impossible to attach several senses to one word, "this would mean a crushing burden on our memory: we would have to possess separate terms for every conceivable subject we might wish to talk about," Ullman (1962, p. 168) further says.

Polysemy by tradition is, in the words of Brown (1986), "the property of an emic expression with more than one meaning" (p.148). For instance, the polysemic noun' bachelor' could mean an unmarried man of any age, a person who has taken the first or lowest degree at a university, a young feudal knight who has followed the banner of another, or a young male fur seal kept from the breeding grounds by older males.

Usually, one of these will fit into a given context, but occasionally, people may need clarification.

HOMONYMY

Homonyms exist when two or more interpretations of 'etic expression' cannot be derived from a common source, that is, from a lexeme having a unique sense. Complete homonyms have the same pronunciation and spelling, e.g., bank (a financial institution) and bank (side of a river).

According to Ullman: "Homonymy is far less common and far less complex than polysemy, though its effects can be just as serious and even more dramatic" (1962,p.176). Homonymy can arise in three ways: phonetic convergence, semantic divergence, and foreign influence.

Homonymy can be divided into homophony and homography. Homophones are words pronounced alike but differ in meaning, origin, or spelling, e.g., knew, new; rite, write, right, right; read (past), red; tale, tail; etc. Homographic words have the exact spelling but are different in meaning, origin, or pronunciation, e.g., record (noun), record (verb), read (present), and read (past).

Discussion on homonymy and polysemy has been largely confined to individual words. However, it is essential to realize that lexical and grammatical ambiguities exist.

LEXICAL POLYSEMY
Prefer =   1. Promote
            2. Like better
'Gentlemen prefer blondes' could be ambiguous in this respect.

Grammatical polysemy
Realisation   Form   Semantics His

Present tense = 1. A momentary happening now. 2. A habitually repeated event, e.g. **"The center-forward, Smith, kicks hard"** is ambiguous in that it might refer to a single event at the time of speaking - reported, say, by a radio commentator - or to a habitual tendency.

LEXICAL HOMONYMY
1. Mole (noun) = A small animal
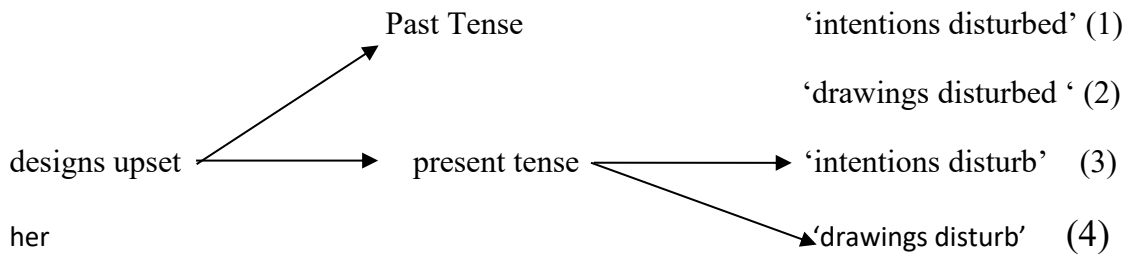2. Mole (noun) = A spot on the skin.
Either meaning is possible in '**I noticed a mole**.'

GRAMMATICAL HOMONYMY
1. **'Moving gates'** as a Modifier + Object construction (gates which move).
2. **'Moving gates'** as a Verbal + Object construction (causing gates to move).
The ambiguity is apparent in "**I like moving gates**.''

As a result of many-one relationships between the levels of semantics, form and realisation, the sentence, "**His designs upset her**", could be assigned four different meanings, as pictured in the following diagram:

Past Tense   'intentions disturbed' (1)

'drawings disturbed ' (2)

designs upset   →   present tense   'intentions disturb'   (3)

her   'drawings disturb'   (4)

The branch in the path of interpretation between realization and form is due to the homonymy of the present tense and past tense of 'upset' (two grammatical functions having the same spoken and written realization); the branching between form and semantics is due to the polysemy of 'designs', which can mean either 'intentions' or 'drawings' in the sentence. Thus, in physical terms, 'the same sentence' can receive four meanings according to context. Lexical ambiguity, it will now be clear, can originate in homonymy, polysemy, or (as in the case of the whole sentence above) a combination of the two.

The context (linguistic or otherwise) does not always permit both readings of an ambiguity to be registered. "**The designs upset her**" would pick out the meaning 'drawings' for 'designs' rather than' intentions.'

THE PROBLEMS OF AMBIGUITY IN SENTENCES

Despite being a common term in grammar, a sentence is not easily defined. It is usually taken as one of the most significant units of grammatical analysis and the largest. The others are clauses, phrases, words, and morphemes.

Attention focuses mainly on its structural characteristics, although these are only sometimes easily distinguishable from those of the clause. Moreover, it is more readily described in its written realization than the spoken, since in speech, sentence boundaries are not easily delimited, and features of form are apt to vary considerably from the norm.

Like many clauses, sentences usually consist of a subject and predicate, but unlike clauses, they can stand on their own as independent units. Traditionally, they are said to contain a complete thought or a distinct proposition.

Sentences have meanings. The literal meaning of a sentence is

determined by the meanings of its component words (or morphemes) and the syntactic rules according to which these elements are combined. Searle (1979) has it that a "sentence may have more than one meaning (ambiguity), or its meaning may be defective or uninterpretable (nonsense)" (p.117).

Lexical ambiguity arises in a sentence when more than one meaning can be assigned to a word. However, structural or syntactic ambiguity arises from using carelessly constructed sentences requiring more formal signals. In the words of Taha (1983), "A grammatical and unambiguous sentence must have some sort of formal signals which help the reader or hearer to recognize the sentence structure" (p.251).

These include function words, word order, inflections, affixes, stress, juncture, etc. According to Fries (1952), structural meaning is signaled by specific and definite devices. He opines that "it is the devices that signal structural meanings which constitute the grammar of a language. The grammar of a language consists of the devices that signal structural meanings" (p.56). When the appropriate structural signals are absent from a sentence, it will be ambiguous. Thus, the sentence can be assigned more than one structural meaning. Hence, the sentence **"Ship sails today"** is ambiguous. The sentence is so because both 'ship' and 'sails' could be nouns or verbs.

However, if the appropriate signals were present, the sentence would be clear, as in 1. **The ship sails today**. 2. **Ship the sails today**. In this case, 'the' signals the structural meaning of 'ship' and 'sails' in the sentences.

Another example is: **"College demands change."** The word 'demands' can be regarded as a noun or a verb. Also, the word 'change' can be interpreted as a noun or a verb. However, if the modal auxiliary 'will' or 'the past tense suffix 'is used with 'demand' as a structural marker, and 'a' or 'some' with 'change,' 'demands and 'change' will be regarded as verb and noun, respectively: 1. **"College will demand a change."** 2. **"College demanded some change."**
Ambiguity may also result from a misplaced participial phrase used adjectivally to modify either preceding noun: **"Okeke pushed his brother reaching for the**

bottle." The ambiguity in this sentence arises due to the participial phrase adjectival 'reaching for the bottle,' which could be regarded as modifying either 'Okeke' or 'brother.' Thus, it is not sure whether 'Okeke' or 'his brother' is the one who is 'reaching for the bottle.' However, the ambiguity will be resolved when the participial phrase adjectival is placed immediately before the noun it modifies, preceding a terminal juncture in speech or a comma in writing: **"Reaching for the bottle, Okeke pushed his brother."**

In a sentence, ambiguity arises if an adverbial modifier follows two verbs and can equally modify either of them. The lexical meanings of sentences of this type, according to Taha (1983), "depend on the direction of the modification; that is, on whether the adverbial would modify one verb or the other in the same sentence" (p.258), as in: **"My father saw the car that was stolen in Ghana."** Here, the ambiguity is that either the car was stolen in Ghana but was seen elsewhere, or the stolen car was seen in Ghana. The sentence will be unambiguous by placing the adverbial modifier close to the verb it modifies, e.g.,

"My father saw in Ghana the car that was stolen."

Another source of ambiguity in a sentence is 'equivocal phrasing.' Here, the individual words are unambiguous. However, their combination can be interpreted in two or more different ways: **"I met several old friends and acquaintances."** The adjective 'old' may refer to friends and 'acquaintances' or only to the former. The context and intonation in the spoken language will clarify the ambiguity of this kind.

In written sentences, ambiguity may further arise as a result of personal pronouns. This happens when a personal pronoun follows two nouns, either of which could be regarded as its reference or antecedent: **"Ogbonna saw a boy talking to his father."** Here, there is no structural signal to help the reader to identify the particular antecedent of the personal pronoun. However, this is disambiguated in speech by the kind of stress received by the personal pronoun: **Ogbonna saw 'a boy' talking to 'his' (boy's) father.** The personal pronoun receives substantial contrastive stress and, thus, refers to the noun immediately preceding it.

On the other hand, if the personal pronoun receives everyday stress, it will refer to the other preceding noun in the sentence:**'Ogbonna' saw a boy talking to 'his' (Ogbonna's) father.**
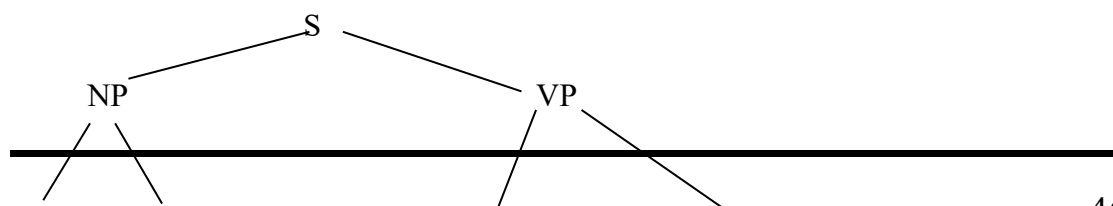
Structural ambiguity can also emanate in sentences from prepositional phrases (PPs), such as **Chidiebere decided on the boat.** This sentence means: 1. **Chidiebere chose the boat.** 2. **Chidiebere made the decision while on the boat.**
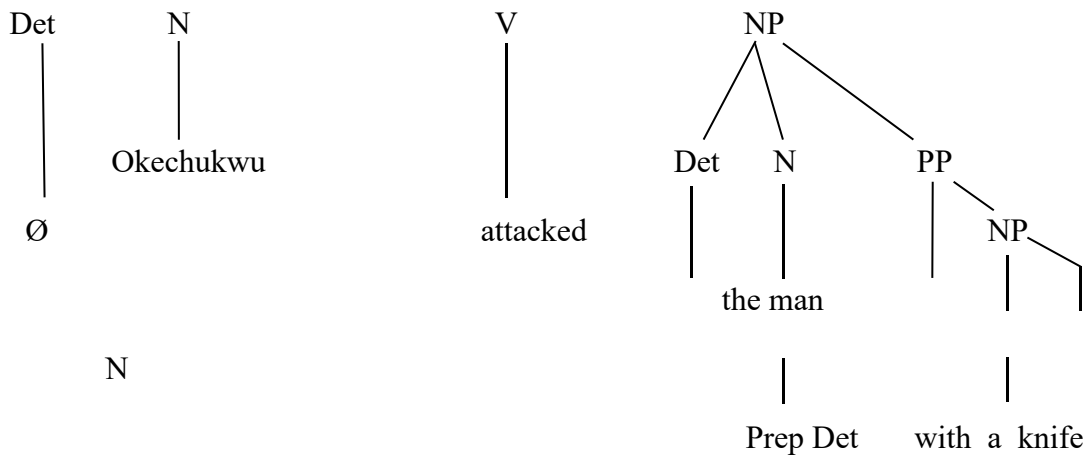
On the (1) reading, the PP is traditionally called a prepositional object, while it has the function of placing adverbial on the (2) reading. Furthermore, the adverbial placement in the (1) reading occurs independently of the selectional requirements of the verb. Instead, it modifies the whole verb phrase (VP). However, the verb selects the prepositional object of (1) just like a direct object.
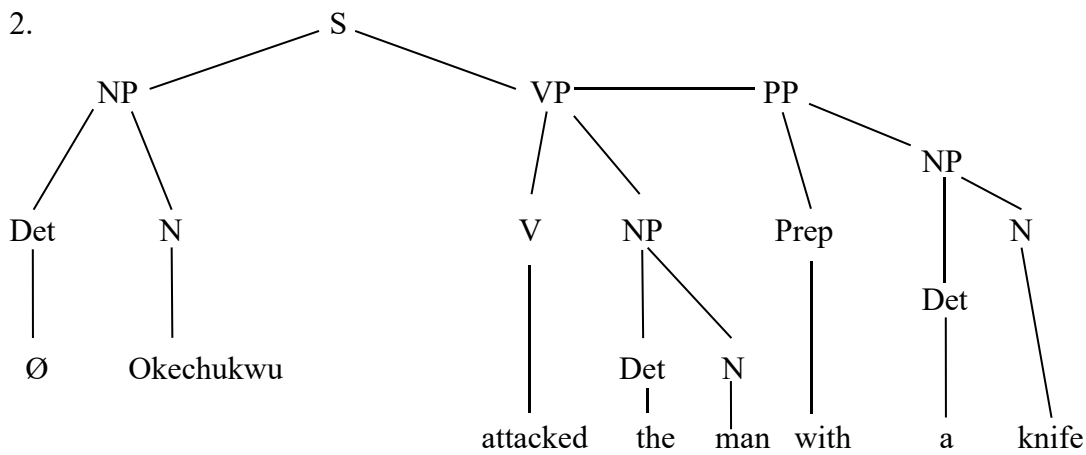
This kind of ambiguity is also apparent in the sentence: **Okechukwu attacked the man with a knife.** The sentence can be interpreted as having the following meanings: 1. **Okechukwu attacked the man who had a knife.** 2. **Okechukwu used a knife to attack the man.**

In (1), 'with a knife' enters into construction with 'the' and 'man,' so 'the man with a knife' forms a single constituent (an NP functioning as an object of 'attacked'). In (2), by contrast, 'attacked the man with a knife' has three immediate constituents: 'attacked' (predicator), 'the man' (object), and 'with a knife' (adjunct). Here, a direct structural relation exists between 'with a knife' and 'attacked,' so 'with a knife' specifies the means of attack. This can be shown in tree diagrams:1.

Det   N   V   NP

| | | / \
| Okechukwu | Det N   PP
| | | | \
Ø   attacked | | NP
                                                      the man | / \
                                                                 | | |
          N   Prep Det with a knife

Okechukwu attacked the man who had a knife.

2.   S
                          /   \
                       NP   VP————PP
                      / \   / \   \
                   Det   N   V   NP   Prep   NP
                    | | | / \ | / \
                    Ø Okechukwu | Det N | Det   N
                                 | | | | |
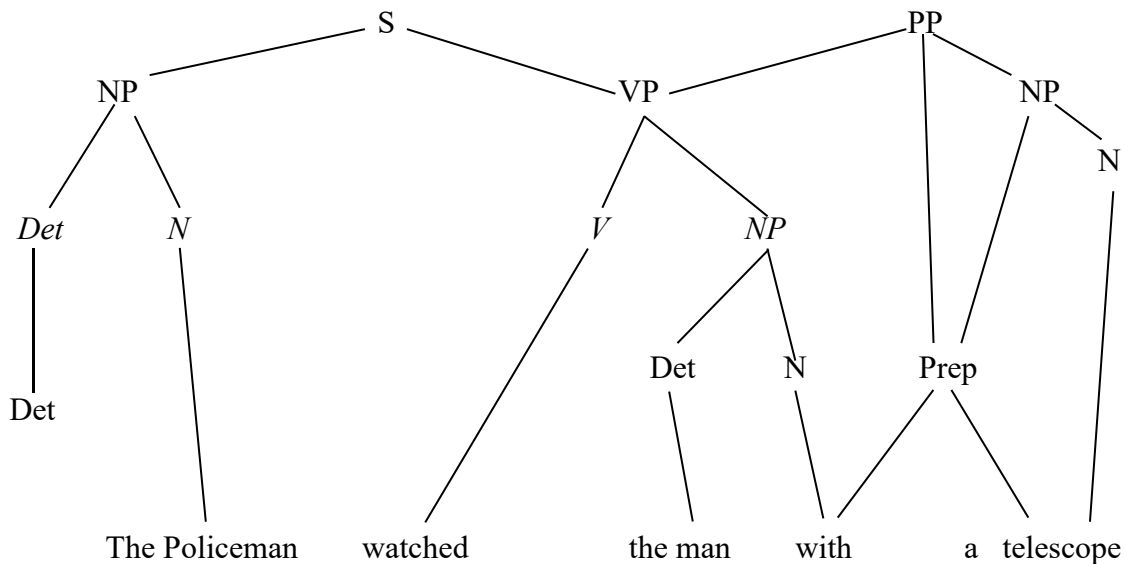   attacked the man with a   knife

Okechukwu used a knife in attacking the man.
The sentence, **"The policeman watched the man with a telescope,"** has
two meanings: 1. **The policeman watched the man who had a telescope.**
2. **The policeman used a telescope in watching the man.** These can be
represented in tree diagrams as well:
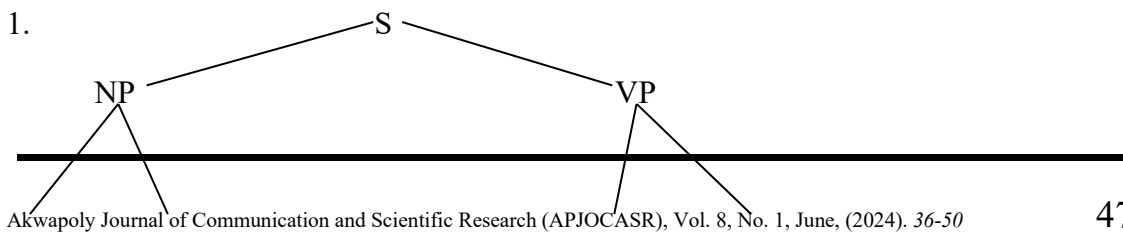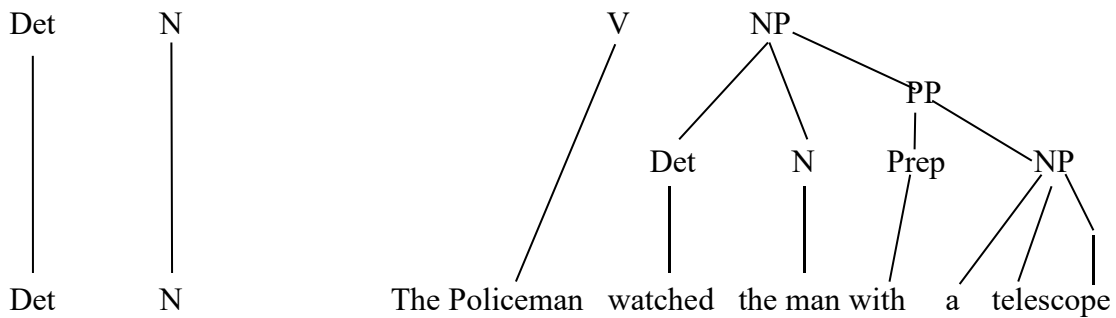
1.   S
                         /   \
                      NP   VP

The policeman watched the man who had a telescope. 2.

The policeman used a telescope in watching the man.

The difference between the PPs is portrayed in a syntactic analysis by showing them with attachments, the different attachments corresponding to different semantic interpretations. The postmodifier relationship of (1) is captured by attaching the PP as a modifier of 'man.' For the instrumental relationship of (2), the PP is attached as a 'daughter' of the VP who 'watched the man.'

CONCLUSION

Ambiguity is found in every language for single words and longer sequences. Lexical ambiguity arises as a result of polysemy and homonymy. If the meanings of polysemous and homonymous lexical items are known, the

disambiguation is carried out successfully; otherwise, the intended or unintended ambiguity passes unnoticed. Sentences that lack structural signals and others whose constituents are unclear bring about syntactic or structural ambiguity. Clear sentence structure can eliminate this ambiguity, and appropriate signals can be added to sentences.

Though the method of knowledge-based approach, grammatical structures analysis, coupled with diagrams adopted in this study, is a further leap among the various methods in the disambiguation of ambiguities in English sentences, more researches are needed for more robust and effective disambiguation of lexical and syntactic ambiguities in natural languages.

REFERENCES

Brown, K. (1986). *Linguistic Meaning*. Vol. 1. London: Routledge and Kegan Paul.

Fellbaum, C. (1998). A Semantic Network of English: The Mother of All WordNets. Computers and the Humanities 32, 209-220 (1998). https://doi.org/10.1023/A:1001181927857

Fries, C. C. (1952). *The Structure of English* as cited in Mark Lester (1976). *Introductory Transformational Grammar of English*, 2nd Edition. London: Holt, Rinehart and Wiston.

Jurafsky, A. & James M. (2009). Speech and Language Processing, 2nd Edition. New Jersey: Prentice-Hall, Inc.

Karpf, A. (1986). The Perception of Ambiguity in a Second Language. *IRAL*, xxiv/2,157.

Manning, C. & Hinrich S. (1999). Foundations of Statistical Natural Language Processing. Cambridge, MA : MIT Press.

Michael, L. (1986). "Automatic sense disambiguation in nepali language" in the Fourth Int'l Conference on Digital Information and Communication Technology and Its Application (DICTAP 2014), Bangkok, Thailand, 2014, pp. 46-50.

Searle, J. R. (1979). *Expression and Meaning*. London: Cambridge University Press.

Sigman, M. & Guillermo A. C. (2002). Global organisation of the Wordnet lexicon 99(3)

1742-1747
htpps.||doi.org|10.1073|pnas.02
2341799

Taha, A. K. (1983). Types of Syntactic Ambiguity in English. *IRAL*, xxi/4,251,258.

Ullman, S. ((1962). *An Introduction to the Science of Meaning*. Oxford: Basil Blackwell and Mott Ltd.

Wales, K. (1989). *A Dictionary of Stylistics*. London: Longman Group Ltd.