Tony Blair Institute for Global Change
Trinity University, Lagos, Nigeria
Harmarth Global Educational Services
FAIR Forward – Artificial Intelligence for All
Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH
Society for Multidisciplinary & Advanced Research Techniques (SMART)

Accra Bespoke Multidisciplinary Innovations Conference (ABMIC)

Grammar-Based Ontology (GBO) Of Signed Words Used In Computer Science Domain

[1]Adejoke O. Olamiti & [2]Victoria Adewale,
Department of Computer Science
[1]University of Ibadan, Ibadan, Nigeria
[2]Crawford University, Igbesa, Ogun State, Nigeria.
E-mails: [1]aolamiti@yahoo.com , [2]victoria.adewale@crawforduniversity.edu.ng
Phones: [1]+2348033671850, [2]+2347066240208

ABSTRACT

Sign language is a visual gesture language that uses hands and arms to convey verbal messages from natural languages using non-manual means such as facial expressions, head movements, hand gestures and body motions. In order to process natural language in the same way that people do, a system must have information about the meanings of the words it uses. In the past, dictionaries were used to provide this information, but machine-readable dictionaries are now commonplace. Dictionary entries, on the other hand, evolved for the benefit of human readers, rather than for machines. This paper implements a Grammar Based Ontology (GBO) of signed words in Computer Science domain for further research work.

Keywords: Computer Science, Domain, Ontology, Sign Language.

Proceedings Citation Format
Adejoke O. Olamiti & Victoria Adewale, (2022): Grammar-Based Ontology (GBO) Of Signed Words Used In Computer Science Domain. Proceedings of the 31st Accra Bespoke Multidisciplinary Innovations Conference. University of Ghana/Academic City University College, Ghana. 1st – 3rd June, 2022. Pp 151-154, www.isteams.net/ghanabespoke2022.
dx.doi.org/10.22624/AIMS/ABMIC2022P16

1. BACKGROUND TO THE STUDY

Sign language is a non-verbal language used by hearing-impaired people to communicate on a daily basis (Schwager & Zeshan, 2008). Information is conveyed visually using a combination of manual and non-manual means of expression. The manual parameters are hand shape, hand pose, hand position, and hand movement. Non-manual parameters include head and body posture, facial expressions, eye movements, and mouth movements. According to research, sign language is a natural human language with a complex composition of lexicons and grammar whose structure is independent of the structure of the surrounding spoken language(s). Sign language includes static and dynamic sign language. However, there is not enough way to capture the dynamic 3D characteristics of sign language utterances on paper.

Therefore, in the absence of integrated multimedia publications with video files, sign language researchers rely more or less on standardized transcription. There are several approaches to sign language translation. The most studied of them are statistical models, rule-based models, data-driven models, and ontology-based approaches (Lozynska & Davydov, 2015).

2. STATEMENT OF PROBLEM

Sign language recognition constitutes a challenging field of research in computer vision. Compared to gesture recognition in controlled environments, recognition of sign language in real world scenarios places significantly higher demands on feature extraction and processing algorithms. Many sign language dataset are distributed in different repositories across the globe and to achieve a robust research, the fusing of the datasets into a single large dataset is essential. In addition to this, sign language have a significant amount of systematic ambiguity or vagueness, and as a result, many signs tend to have rather general meanings that are narrowed down by the context of the utterance, and because many grammatical categories can remain unmarked, a sequence of signs can be difficult to categorize structurally, even if its semantic meaning is entirely clear.

3. OBJECTIVE

This study focuses developing a robust American Sign Language (ASL) dataset for basic Computer science signs from sign language databanks and other local sources. The datasets will be structured using Grammar Based Ontology (GBO) of signed words for further research work.

4. METHODOLOGY

4.1 The Research Design

The ASL sign data used in this study is derived from the MS-ASL dataset (Joze et al., 2019), which permits the examination of generalization to unseen individuals (signer-independent test) in a realistic setting with more than 200 signers. This dataset contains a large class count of 1000 signs recorded under challenging and unconstrained real-world recording conditions, as opposed to the previous work's focus on limited vocabulary tasks. To create a robust signbank for static and non-static signs, additional data sources will be added.
Ontology is a set of concepts and categories in a given subject area or domain that demonstrates their properties and relationships. Classes, attributes, relationships, and instances are regarded as basic elements in ontology, which is an explicit description of concepts. Putting the Computer Science domain into consideration, an Ontology (O) can be represented as the tuple O = (C, R, I, P), where; C is a set of classes (domain concepts), R is a set of relationships between classes in C (matches predicates to verb concepts), I is a set of class instances (concrete objects), and P is a set of predicates. Natural language is defined by grammatical rules, and these rules are applicable in creating the ontology for the chosen domain.

5. DATA PRESENTATION

In eliciting the requirements to create an grammar-based ontology of frequently used words from wordnet in the computer science domain, the following stages are considered:

1. Look up each word in the sentence's definition.
2. Analyze the grammatical characteristics of each word by creating base forms.
3. Use hypernyms and hyponyms to describe the meaning of each word.
4. To each verb in the sentence, add all expressions.
5. Use a Parser to parse the sentence.

The Synset (Lualdi et al., 2021) function in python NLTK (Natural Language Tool Kit) was used to look up words in the WordNet database, with base forms generated from word_forms with their grammatical characteristics. e.g. get_word_forms('computer')

- {'n': {'computation', 'computers', 'computerizations', 'computerization', 'computer', 'computations'},
- 'a': {'computational'},
- 'v': {'computerizing', 'compute', 'computerise', 'computing', 'computes', 'computerizes', 'computed', 'computerize', 'computerized'},
- 'r': {'computationally'}} where the Part of Speech (POS) sysnets [n, a, v, r] represents n*oun, adjective, verb* and *adverb* respectively. To increase the likelihood that two words' meanings would be associated, hypernyms was found to be a useful tool in this process, while hyponym forms the subset of the entity in the class. For instance, the entity computer: **Synset name:** computer.n.01 was found to have the following hypernyms/hyponyms;

Synset specific term: [Synset('assembly.n.01'),
Synset('bagger.n.02'), Synset('calculator.n.02'),
Synset('calender.n.01'), Synset('cash_machine.n.01'),
Synset('comber.n.03'), Synset('computer.n.01'),
Synset('concrete_mixer.n.01'), Synset('corker.n.02'),
Synset('cotton_gin.n.01'), Synset('decoder.n.02'),
Synset('farm_machine.n.01'), Synset('franking_machine.n.01'),
Synset('hop-picker.n.01'), Synset('machine_tool.n.01'),
Synset('machinery.n.01'), Synset('milking_machine.n.01'),
Synset('motor.n.01'), Synset('pavior.n.01'),
Synset('perpetual_motion_machine.n.01'), Synset('pile_driver.n.01'),
Synset('power_shovel.n.01'), Synset('power_tool.n.01'), Synset('press.n.03'),
Synset('press.n.07'), Synset('printer.n.03'), Synset('record_player.n.01'),
Synset('riveting_machine.n.01'), Synset('self-feeder.n.01'),
Synset('simulator.n.01'), Synset('slicer.n.02'),
Synset('slot_machine.n.01'), Synset('snow_thrower.n.01'),
Synset('sorter.n.02'), Synset('stamp.n.07'),
Synset('staple_gun.n.01'), Synset('stapler.n.01'),
Synset('textile_machine.n.01'), Synset('time_machine.n.01'),
Synset('trimmer.n.02'), Synset('workhorse.n.01'), Synset('zamboni.n.01')]

Expressions were added to the words to form sentences using CFG, thereafter the NLTK parser was used to generate the parse tree. E.g. Grammar (G): Grammar with 14 productions (start state = S) S -> NP VP

```
PP -> P NP
NP -> D Noun
NP -> NP PP
VP -> V NP
VP -> VP PP
D-> 'a'
D -> 'the'
Noun -> 'Lecturer'
Noun -> 'student'
V -> 'teaches'
V -> 'learns'
Prep -> 'from'
Prep -> 'to'
```

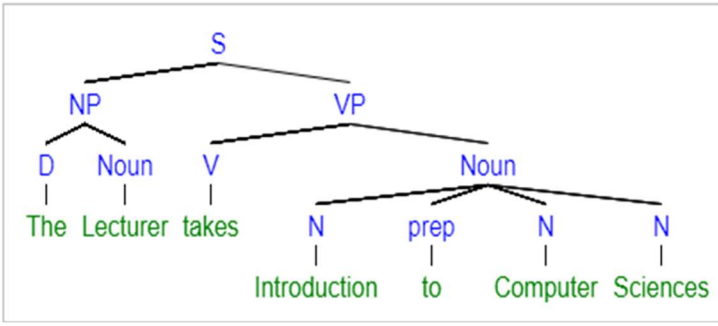

Figure 1: Sentence Formation

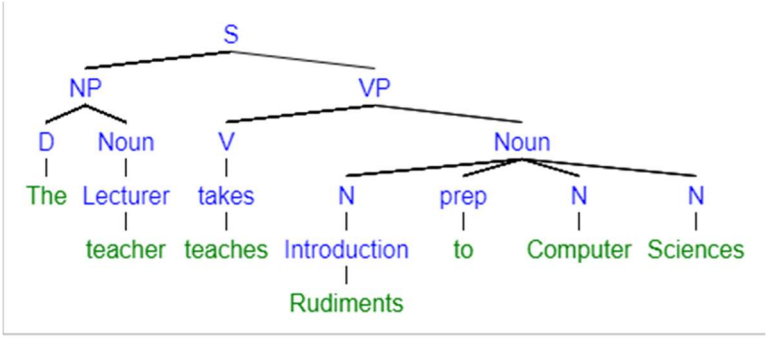

Figure 2: Synonymous Grammatical construction

6. CONCLUDING REMARKS

The NLTK python environment. (Bird et al., 2009) was used to create the domain-specific GBO. Using this method, it was possible to edit ontology files in a Python environment with intelligent code completion. Grammar -Based Ontology for signed words can be extended to other domains for future research.

7. CONTRIBUTIONS TO KNOWLEDGE

The introduction of this study is to provide standardized grammar structure in the usage of natural language in sign language. The success of the research will enhance formal learning of English language by the hearing impaired and also facilitate better understanding in communication with the larger society.

REFERENCES

1. Bird, Steven, Loper, E., & Klein, E. (2009). *Natural Language Processing with Python.* O'Reilly Inc.
2. Joze, V., Reza, H., Koller, & Oscar. (2019). *MS-ASL: A Large-Scale Data Set and Benchmark for Understanding American Sign Language*. he British Machine Vision Conference (BMVC).
3. Lozynska, O., & Davydov, M. (2015). Information technology for Ukrainian Sign Language translation based on ontologies. *ECONTECHMOD. AN INTERNATIONAL QUARTERLY JOURNAL* –, *04*(02), 13-18.
4. Lualdi, C. P., Wright, E., Hudson, J., Caselli, N. K., & Fellbaum, C. (2021). Implementing ASLNet V1.0: Progress and Plans.
5. Schwager, W., & Zeshan, U. (2008). Word classes in sign languages. *Studies in Language*, 509-545. doi 10.1075/sl.32.3.03sch