Lateral Inversions, Word Form/Order, Unnamed Grammatical Entities and Ambiguities in the Constituency Parsing and Annotation of the Igala Syntax through the English Language

Mahmud Mohammed Momoh
Prince Abubakar Audu University
Anyigba, Kogi state, Nigeria
Mahmoodmohammed19@yahoo.com

Abstract

The aim of this paper is to expose the structural form of the Igala language and the inherent complexity related to the translation of the language to a second language vis-à-vis the English language through a configurational probing of its word order, lateral inversions, and unnamed grammatical entities in relation to parsing and annotation in computing. While this study finds out that there is a preponderance of a linguistic typology with subject-verb-object word order and the near total absence of preposition in the speech composition of the Igala language, this fact has not been taken as a serious subject for intellectual consideration. In this study, the abstruseness or incongruity associated with interpreting the Igala syntax through part-of-speech (POS) tagging in relation to its word order, lateral inversion of some phrases, and unnamed grammatical entities (i.e. preposition) in its speech processing into English shall be exposed. Thus, generating a comprehension model for automotive identification, application and/or conversion of these structural forms to the English language shall be the focus of this paper

Keywords: lateral inversion, word order, unnamed grammatical entities, parsing, annotation, Igala, English

1. Introduction

Past works on translation of the Igala language to a second language have focused on the effectiveness of using the English language combined with Igala in teaching in primary schools (Achor and Akor, 2015), evolvement of a modeled language processor that can accept as input Noun Phrases in English language and translate these to Igala (Ayegba, Osuagwu, and Okechukwu, 2014), example acquisition (alignment), matching and recombination (Joshua, Ayegba, and Ojochegbe, 2020), syntactic interference (Attabor, 2019), and contrastive analysis on the use of conjunction (Abraham, 2017). It is worth noting that, no special focus has been placed on the unnamed grammatical entities, word ordering, and the parsing and annotation of inherent syntactic structures. This is notwithstanding the fact that, variations in grammatical rules, word forms and syntactic sequences could be a source of ambiguity and difficulty in translation and comprehension from Igala vis-à-vis the English language by both the machines and the physical learners. This sort of ambiguity has been proven in a more typical sense in regard to translation from a pro-drop language like Japanese or Korean to a non-pro-drop equivalent like English (Wang, Tu, Zhang, 2017). Although, I found out that despite the fact that the Igala language like the English language (see Dryer),[1] French (Bonami, Godard, and Marandin, 1999), Italian (Brunato and Dell-Orletta, 2017), (Namboodiripad, Kim, and Kim, 2017), anchors

mainly on a single word order (i.e. subject-verb-object (SVO)), there was still translational ambiguity in implementing an accurate syntactic parsing and annotation for the two languages. Ambiguity in translation from Igala to the English language aside, this sort of mismatch in parsing and annotation could be more serious when carrying out machine-based translation (MT) between Igala and the other languages with contrastive or differential word order such as Korean (Minhui, and Emily, 2015) which uses the postpositional speech form (Mun and Desagulier, 2022) or as in Afaan Oromo (Meshesha, and Solomon, 2018), verb-object-subject (VOS) order as in Malagasy (Ileana, and Postdam, 2024), verb-subject-object (VSO) order as in Welsh (Borsley, Tallerman, and Willis, 2007) or Old Irish (McCone, 1997), or object-verb-subject (OVS) word order as with the not so popular Cariban language; Hixkaryana (Kalin, 2014), in Brazil.

Unlike the observation by Minhui, and Emily (2015) and Namboodiripad, Kim, and Kim (2017) for the Korean language as well as another observation by Fransen (2020) for Old Irish concerning the inherence of multiple word ordering format, I found that the Igala language dwells mainly on a single word order, i.e. Subject-verb-object, as in the phrase; *ū l'ōpā* ≡ "I chewed groundnut" which has the same grammatical approximation in meaning and word sequence with English. However, a notable challenge bedeviling the parsing and annotation of the Igala syntax, most especially with its conversion to English is that of the lateral inversion of some syntactic forms and phrases as

---

I did observe in respect to this; "ȯkwō wē wā", which is sequentially or literally; "grandparent your came", but actually; "your grandparent came" in the English language.

More also, despite the translation complexity that arises from translation and language teaching when a given part of speech existing in one language does not exist in a corresponding language, from my findings, there are no clearly defined prepositions (which together with postpositions was sometimes referred to as the non-lexical heads of phrases) (Frazier, 1980) in Igala, and thus resulting in incomplete sequential word outlays, vagueness or obscuration of the basic order typology of natural languages and unclear understanding due to this lack of word alternatives during parsing, annotation and general translation as Boquist (2009)[2] did also observed. In this paper, Igala syntactic forms lacking or not containing prepositions would be parsed through parse trees and the corresponding annotations would be converted to the English language as a way of exposing gaps in correspondence and determining the accuracy of translation.

Following the successes of Warren Weaver in the 1950s and the successes that have been recorded in machine translation thereafter – especially in the aspect of part-of-speech tagging in machine translation as Guidivada, and Arbabifard (2018) did rightly observed, I was able to parse and annotate the syntactic structure of the Igala through the English language. Acting upon the suggestion of Guidivada, and Arbabifard (2018) and Jurafsky and Martins (2009), a *transfer-based approach* which uses a three step process was adopted in the segmental structuring of this paper. First, some syntactic analysis (e.g., building a parse tree) is performed on the source text. Second, the syntactic structure is converted (i.e. transferred) into a corresponding structure in the target language. Finally, output is generated from the syntactic structure of the target language. The orthographic frame used in this work as well as the rule of elision expressed in subsection 2.2 conforms to the form adopted in Momoh (2023) and the video on.[3] Furthermore, the triple bar symbol was used to represent equivalence in translation from English to Igala while the approximately equal to symbol ≅ was used to express syntactic isomorphism in translation of syntactic form having differences in lateral sequence of words between Igala and English but the same meaning upon translation to the English language. Using segmented Treebank, six trees-bearing graphs were

designed where figures I, II and III contains expressions in the Igala language while IV, V, and VI deals with the English language. Thereafter, the output of the parsing and annotation done was used in demonstrating the structural form of the Machine-based translator being proposed in this paper.

2. Syntactic Analysis, Parsing and Annotation in Igala

In a holistic sense, the Igala language mainly uses the subject-verb-object word order.

2.1 Syntax Parsing and Annotation of the Igala Inverse Possessive Determiners using English

Before designing a parse tree to demonstrate this form of word sequencing in the Igala syntax, the three pronouns; *mā* (their), *mī* (my), and, *nwū* (his/her/it) are considered in respect of their syntactic applications to the Igala phrases demonstrated in the three forms.

Ōmā mā kwū ōrōkā ōnālē ≡ "child their died afternoon yesterday" ≅ "their child died yesterday afternoon";

Īyē mī wā ≡ "mother my came" ≅ "my mother came"; and

ȯwȯ nwū dē ≡ "goat his/her be goat" ≅ "this is his/her".

I then did the parsing using the first of the three possessives (i.e. *mā*). The first sentence – 'Ōmā mā kwū ōrōkā ōnālē' was represented by the parse tree in Figure I;
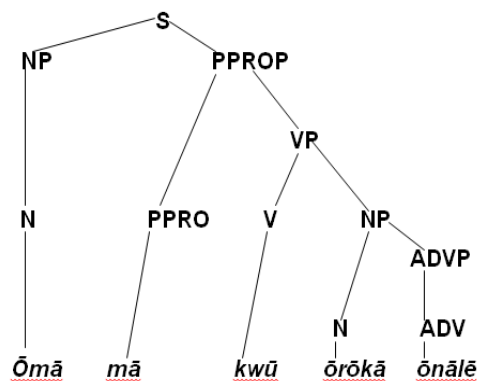

Figure I: Parsing of inverse syntactic determers in Igala.

The next step which was in line with the model proposed in[4] (a system used by the Penn

[2]https://digitalcommons.liberty.edu/cgi/viewcontent.cgi?article=1106&context=honors
[3]https://doi.org/10.48448/e0np-e385
[4]https://www.lancaster.ac.uk/fss/courses/ling/corpus/Corpus2/2PARSE.HTM#:~:text=This%20term%20alludes%20to%20the,article%2C%20P%3Dpreposition.)

Treebank project) (Marcus, Kim, and Marcinkiewicz et al 1994; Santorini, 1990) was to provide a bracket-based morphsyntactic annotation using underscore character (_) in the form of part of speech tags and the use of square brackets annotated at the beginning and the end with the phrase type [s......] as thus:

[S [NP *Ōmā*_NP1 NP] [PPROP *mā*_ PPROP [VP *kwū*_VVD [NP *ōrōkā*_NN1 NP] ADVP_ *ōnālē*] S]

This was also written alternatively as;

[S
   [NP *Ōmā* NP]
   [PPROP *mā*
      [VP *kwū*
         [NP *ōrōkā* NP]
       [ADVP *ōnālē*]
S]

2.2 Subject-Verb-Object Word Order Parsing and Annotation in Igala

The form of word ordering used in this sub-section follows the same pattern as in the English language. Some phrases and sentences use subjective personal pronouns 'I', 'you' (both in the singular and in the plural form), he/she/it, we, they, and who. While I provided sentences bearing these forms of subjective personal pronouns with respect to these being objects of sentences, both the Treebank and the annotation with respect to this form of word order was done using the subjective form of proper nouns and common nouns by which I provided only one example.

Subjective personal pronouns as subject of the sentence examples:

òmī k'ōmāgòlò (I plucked mango) which is simplified albeit ⃠ (unconventionally prohibited in writings) as *òmī kā ōmāgòlò* (*òmī* + *kā* + *ōmāgòlò*) = (me + plugged + mango);

ē/me wė ālū (you (singular)/you (plural) + shut + mouth), translated literally as (*ē/me* + *wė* + *ālū*) = (you/you + shut + mouth);

ī w'ūnyī (she/he/it came home) which is simplified albeit ⃠ as *ī wā ūnyī* (*ī* + *wā* + *ūnyī*) = (she/he/it + came + home);

āwā d'ūnyi (we be home/we are home) which is simplified albeit ⃠ as *āwā* + *dē* + *ūnyi* (we + be + home);

āmā d'ōbē (they took the knife) which is simplified albeit ⃠ as *āmā* + *dū* + *ōbē* (they + took + knife); and,

ēnē k'āfē? (Who took the cloth?) Which is simplified albeit ⃠ as *ēnē* + *kò* + *āfē?* (Who + took + cloth?).

The next step that I took was to frame a sentence with a proper noun as the subject of the sentence and a common noun as the sobject of the sentence as was done in https://www.lancaster.ac.uk/fss/courses/. This was done because the form of word ordering considered in this subsection follows the same word order as English which was the language annotated in https://www.lancaster.ac.uk/fss/courses/. The sentence '*Ūgbédė gw'òjī òdė kā*' which translates as (Ugbede sat on a stool) is the example used. I found out that should the so called prevailing rule on 'conventionality' which adopts apostrophe (as in the word *gw'òjī* above) to fuse two words to one should win through or remain consolidated with respect to machine translation (MT), morphosyntactic annotation of texts becomes complicated. As in the case of the so called 'phrase' *gw'òjī* which can be split to the two separate words gwu which means 'sit' in English and *òjī* which also means 'head' in English but also used to mean 'on', 'above' or 'over' in respect to the dual fusion '*l'òjī*' (pass head) in a more figurative sense (or 'went over' in an actual sense) because of a want of alternative word for expressing the word 'on'.

Thus, in following with the call for the "expansion of contracted forms of multiple words, so that all the words have well defined grammatical categories",[5] in annotating the sentence, 'Ugbede sat on a stool', I used the so called 'unconventional' form of writing the sentence '*Ūgbédė gwū òjī òdė kā*' rather than '*Ūgbédé gw'òjī òdė*'. The reason being that the former (i.e. *Ūgbédė gwū òjī òdė kā*') is amenable to parsing and annotation as it is in line with the Penn Treebank Project while the latter (i.e. '*Ūgbédė gw'òjī òdė kā*') is not. Here too, the indefinite article 'a' was substituted with the indefinite pronoun '*kā*' which translates in English as 'one'. Although, I found out that articles are classified as separate part of speech in their own right but since they are also considered as a kind of determiners and the word 'one' can be used as a determiner, reference to the word 'one' as used in the sentence is classified as an article and treated as such in the Treebank presented in Figure II.

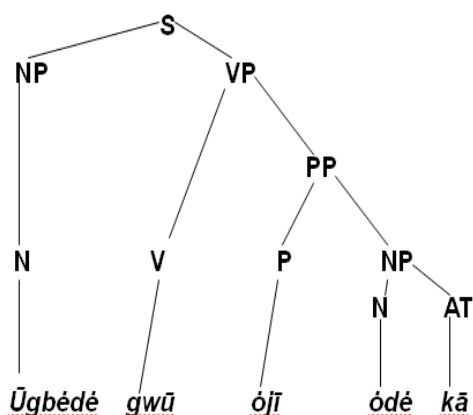[5] https://www.cs.rochester.edu/u/brown/242/assts/termprojs/micha/docs/parser.html

Figure II: Subject-verb-object word order parse tree in Igala.

Next, a second bracket-based morphsyntactic annotation using underscore character (_) in the form of part of speech tags and the use of square brackets annotated at the beginning and the end with the phrase type [s......] was provided in respect of Figure II as thus;

[S [NP *Ūgbédé*_NP1 NP] [VP *gwū*_VVD [NP *ójī*_ II [NP *ódé*_NN1 *kā*_AT1 NP] NP] VP] S]

This was also written alternatively as;

[S
  [NP *Ūgbédé* NP]
   [VP *gwū*
     [NP *ójī*
        [NP *ódé* *kā* NP]
       NP]
     VP]
S]

2.3 Unnamed Prepositional Entities and Constituency Parsing and Annotation in Igala

In this subsection, I made reference to the 9th Edition of the *Oxford Advanced Learner's Dictionary of English* in which the word preposition was defined as - "a word or group of words such as *in, from, to, out of*, and *on behalf of*, used before a noun or pronoun to show place, position, time or method" (Hornby, 2015). Added to this five (in, from, to, out of, and on behalf of) examples of preposition above were eleven more examples culled from,[6] that included; "beneath," "beside", "between", "in front of", "inside", "near", "off", "through", "toward", "under", and "within". Although, there is the argument adduced by Ilori (2015) to support his claim that there are named prepositions as part-of-speech in Igala for which he went as far as counteracting the claims adduced by other writers like Atadoga (2011) and

Ikani (2011) regarding the use of body parts as prepositions but through a careful assessment of the prepositional forms in English pointed out from Hornby (2015) above, I found out that, in truth, the syntactic form of the Igala language does not contain preposition in a more specified sense of the word. In this subsection, I shall cite one example drawn from Ilori (2015)'s abstract where he regarded the word '*tū*' as contained in the phrase '*tú unyí un*' in which he probably meant to say that the word '*tū*' specifically implies the English word 'to' when in reality '*tū*' meant 'unpack' or 'unfasten' while '*tū*' in respect to the preposition has no syntactic base and only exists when its 'root' (the 't') is tied with the word '*ūnyī*' (house or home in English) as in the form '*tūnyī*' as I did pointed out in subsection 2.2 with respect to the word(s) '*gw'ójī*' or '*gwū ójī*' and how this form of dual-word contraction through elision or as a matter of convenience could be a source of ambiguity or encumbrance to word encoding in the design of parse trees and annotation.

In the next lines, I shall try to demonstrate how prepositions are unnamed entities in the syntactic framing of sentences in Igala using the five examples of prepositions offered by (Hornby, 2015) above.

With respect to 'in'; "*ódūdū à wa*" (morning + we + come) which actually translates as "in the morning we shall come" or "we come in the morning";

In respect to 'from'; "*ōmō ī kwó*' (there + he/she/it + left) which actually means ("he/she/it came from there" in English). In a sense, the verb 'left' is used instead of 'from' in Igala grammar;

Reference has already been made to the word 'to' above so there is no point adding extra expression to that here;

With respect to 'out of'; "*éfū mā ī kwó*" (belly + them + it + came) which actually meant ("out of them it came") in English;

With respect to 'on behalf'; "*t'ódū mī*" (t + name + me) which actually means ("because of me" in English).

Thus, using the first sentence in respect to the word 'in', a simple parse tree with its annotation was provided to shed more light on this. Figure III has the parse tree.
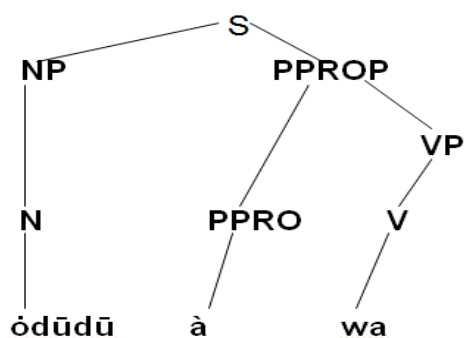
Figure III: Parse tree expressing unnamed preposition in Igala.

I then framed an annotation for the syntactic form of the parse tree in Figure III as thus;

[S
   [NP ódūdū NP]
   [PPRO à
     [VP wa
      NP]
    VP]
S]

3. Conversion of the Igala Syntactic Form to English

In this section, I converted the parse tree/annotation in the preceding section into English in the form of a translation.

3.1 Conversion of the Igala Possessive Determiners to English

I found that, owing to the lateral inversion of syntaxes, there is a disproportionate incongruity in converting syntactic forms in Igala to English in a figurative sense as demonstrated below. What I did was to reverse the phrases used in the second section above from English to Igala.

Thus, the three forms of pronouns; their (*mā*), my (*mī*), and, his/her (*nwū*) are considered in respect of their syntactic applications in the English phrases given in the example below.

"Child their died afternoon yesterday" ≡ *Ōmā mā kwū ōrōkā ōnālē.*

I then created a parse tree representing this word order in English as thus;
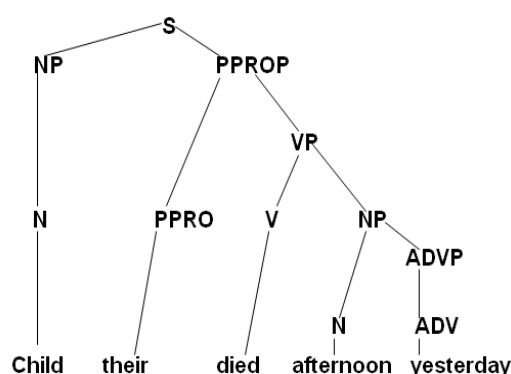

Figure IV: Parse tree showing conversion of possessive determiners from Igala to English.

Figure IV was annotated a thus;

[S [NP Child_NP1 NP] [PPROP their_ PPROP [VP died_VVD [NP afternoon_NN1 NP] ADVP_ yesterday] S]

This was also written alternatively as;

[S
   [NP Child NP]
   [PPROP their
     [VP died
      [NP afternoon NP]
     [ADVP yesterday]
S]

3.2 Conversion of Subject-Verb-Object Word Order from Igala to English

Notwithstanding the fact that the English subject-verb-object word order also exists in Igala, getting an accurate translation for English to Igala proved a little bit problematic as shown in Figure V.
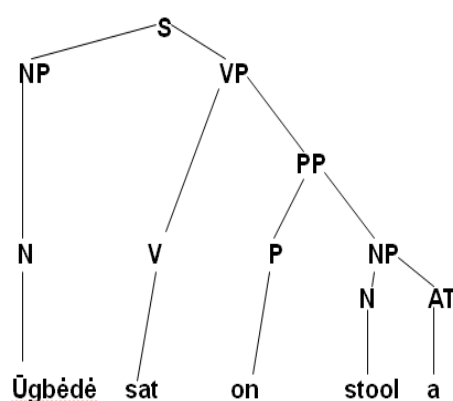

Figure V: Parse tree showing the conversion of subject-verb-object word order from Igala to English.

A bracket-based syntactic annotation for Figure V was given below;

[S [NP Ugbede_NP1 NP] [VP sat_VVD [PP on_ II [NP stool_NN1 a_AT1 NP] PP] VP] S]

This was also written alternatively as;

```
[S
    [NP Ugbede NP]
    [VP sat
        [PP on
            [NP stool a NP]
        PP]
    VP]
S]
```

3.3 Conversion of Igala Syntactic Forms with Unnamed Prepositional Entities through Parsing and Annotation in English

Here, what I did was to copy the phrase used for the parsing and annotation of the sentence bearing the unnamed preposition in subsection 2.3 (i.e. "*ódūdū à wa*" - (morning + we + come) which was used to build a corresponding parsing and annotation in English. This was represented in the constituency parsing on Figure VI and the annotation that comes below it.
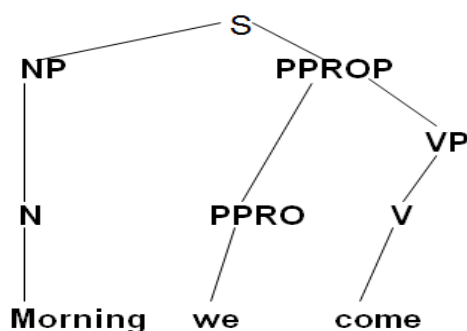

Figure VI: Parse tree showing the conversion of unnamed prepositional entities from Igala to English.

The syntactic form of the phrase on Figure VI was annotated as thus;

```
[S
    [NP ódūdū NP]
    [PPRO à
        [VP wa
            NP]
    VP]
S]
```

4. Output Generated from the Grammatical Structure of the Igala Language

From findings in this work, it becomes clear that annotation of the syntactic form of the Igala language would remain a herculean task. The fact that certain words are being conjoined arbitrarily as one through the use of apostrophe makes it hard for words to maintain their original form during sentence composition, making it hard for their annotation and translation to a second language. The implication of this is the presence of mixed signal while trying to convert syntactic form from the Igala language to English in a more specific sense.

Thus, in using the apostrophe for conjoining two words as one which is currently the case among most writers of the Igala language in which case, the phrase 'leave there' becomes '*kw'ōmō*' /kwomo/ rather than '*kwò ōmō*' /kweu omo/ and 'put there' becomes, '*t'ōmō*' rather than '*tō ōmō*' - a practice done as a way of endearing fluency in conversation (Momoh 2023). Being an isolating language agglutinating inflectional morphemes with more than one unit of meanings denoted by separate part-of-speech, how to encode the specific word and them along their individual grammatical unit during parsing and annotation for a working machine-based translation becomes difficult. While most words are classed as having a 1:1 morpheme per word ratio, others like '*k'ōmō*' /komo/ (hit there), '*g'ōmō*' /gomo/ (look there), have a 2:1 morpheme per word ratio that is similar to the explanation provided in respect to Russian by Comer (2021). Following from this fact, this writer found that the syntactic codes for part-of-speech (POS) parsing and annotation proposed in the Penn Treebank Project are insufficient for the parsing and annotation of the Igala language. Whether to use special identifiers such as the plus sign (+) or the slash sign (/) in expressing agglutination, i.e. to express the parsed form of *gomo* (look there) as VP+ADVP or V+ADV and VP/ADVP or V/ADV on the vertical dashes or whether to have words like '*k'ōmō*' /komo/ (hit there), '*g'ōmō*' /gomo/ (look there) written without the use of the eliciting mark expressed by the application of the apostrophe remains an issue of concern. Although the use of the + (plus sign) as suggested here comes with a different mode of application, but this comes close to the same indicator used for analyzing contraction as PPSM+BEM in the pioneering Brown Corpus (Marcus, Santorini, and Marcinkiewicz, 1993).

With respect to the parse tree and annotation of the inversed possessive determiners, I found just a 20 per cent mean correlation in the sequence of word order in the syntactic translation from Igala to English and vice versa, with the result that, out of the five words used apiece, only the median word '*kwū*' and 'died' maintained consistency in the sequence of word arrangement as shown in the third vertical dashes on the two figures (I and IV) representing the parse trees and also on their individual annotations. There was also an attendant displacement of four (*Ōmā, mā, ōrōkā,* and *ōnālē*) of the five Igala words and four (child, their, afternoon, and yesterday) of the five English words upon conversion from Igala to English. Following from this fact, I found an 80 percentage point to this end. More also, owing to the nonexistence of preposition in the Igala word forms, the word '*ójī*' (head) – but could as well be

translated as 'thief' in English and which was represented as a noun on the third vertical dash of the parse tree of Figure II was replaced with the word 'on' – a preposition, upon conversion to English on Figure V. The implication of this is a noun + noun sequence in the syntactic order of the phrase 'ọ́jī ọ́dẹ́' as was also done by Ilori (2015) on page 146 of his paper.

It therefore implies that the word 'ọ́jī' in a more figurative sense would have to be recognized as 'on' during word conversion through parsing and annotation in English and in which case, both the parse tree and the annotation of the phrase 'Ūgbẹ́dẹ́ gwọ̀jī ọ́dẹ́ kā' would have to be redrawn in line with the Brown Corpus format as thus;
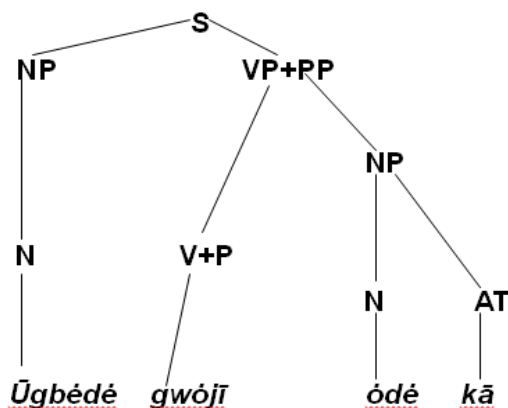

Figure VII: Parse tree showing the splitting of gwọ̀jī into gwū and ọ́jī.

[S [NP *Ūgbẹ́dẹ́*_NP1 NP] [VP+PP *gwọ̀jī*_ II [NP *ọ́dẹ́*_NN1 *kā*_AT1 NP] VP+PP] S]

This was also written alternatively as;

[S
    [NP *Ūgbẹ́dẹ́* NP]
     [VP+PP *gwọ̀jī*
            [NP *ọ́dẹ́ kā* NP]
         NP]
      VP+PP]
S]

More also, the Treebank and the syntax annotation with respect to table II and IV shows an 80 per cent correlation in word sequence with the result that, while there is no word elision as it is in respect to figures I and IV, the words 'a' and 'stool' were however inversed laterally from the ordering sequence they exist in the Igala syntactic structure in so that, the syntactic or phraseological form of 'a stool' in the English language became reversed as 'stool a' or 'stool one' ⟺ *ọ́dẹ́ kā*. Rising from this fact that none of the five indicators represented in figures I through VI was unnamed in the two languages, a 60 per cent translation accuracy using the subject-verb-object for Igala and English was arrived at. I found that from the

five syntactic variables exemplified by the five vertical dashes on the parse trees on figures IV and V there was an accuracy in word frequency of 3 > 2 and an error of 2 < 3. The implication of this is that there was a 40 per cent error towards this end.

Through figures III and VI we also noticed two kinds of errors or inaccuracy in translation from the sequence of word on the figures as was also apparent in the flow pattern of the annotation of the content of the parse trees on both figures. There were cases of lateral inversion in translation from Igala to English.

Figure VIII is a two-way crawling translator that can also be a word-to-meaning finder through the pipes connecting A1 and A2 and B1 and B2. Input is received via either side of the translator with the blue colour representing channels for the flow and transmission of words in Igala while the orange colour boxes represent the English equivalent. The vertical rectangle in either portion of the three boxes coded I (input) is the transformer which is connected to eight word banks representing the eight parts of speech in which the corpuses would be fed. A1 and B1 are word receivers while A2 and B2 are parsers but can also function alternatively as input and output processor if the machine is commanded to find word and meaning in the given language.

A1 and B1 are machine-based parsers and/or annotators that decode questions transmitted from the affected I boxes of the source language(s) while A2 and B2 are parse and/or annotation converter into the target language. With the syntax moved through the various I boxes, these are sent into the transformer.

From the eight (8) boxes fed lemmas or words according to the given part of speech category of each lemma, i.e. boxes attached to C1 and D1, these lemmas and their meanings as stored in each of the eight boxes are connected directly through the eight pipes linking the eight boxes to the word receivers attached to the individual transformer.
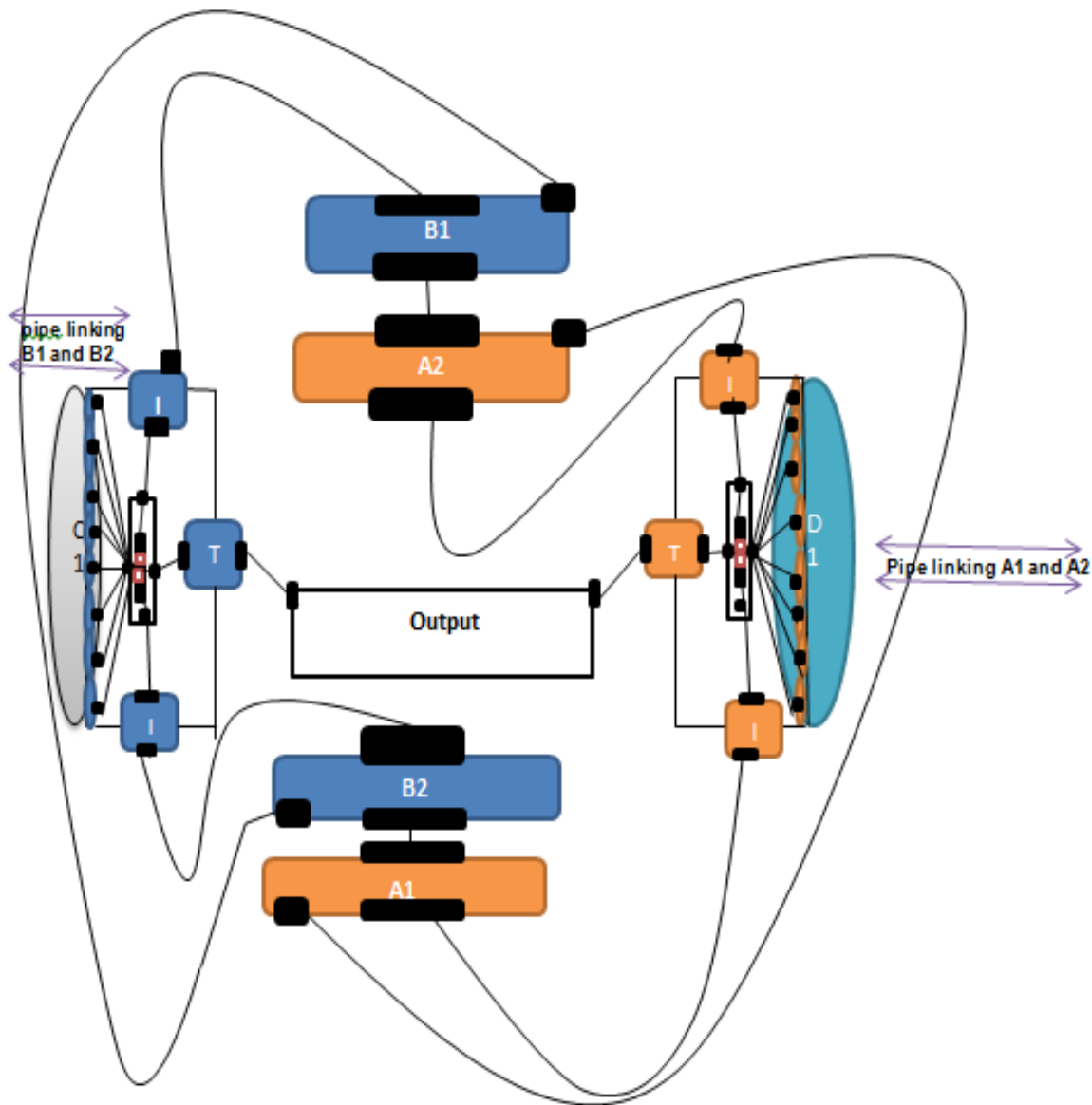
Figure VIII: Bidirectional model for the design of a Machine Translator of English to Igala and from Igala to English.

When syntax received by A2 and B2 are parsed/annotated through the converter to be built into A2 and B2, they are moved through the next I channel on the pipeline to the transformers on either side of the model. Rather than translating the syntax directly, the transformer, using scanners connected to it, finds words equivalents in the part-of-speech boxes, using the scanners and word receivers attached to the individual transformer. Through the word receiver, the individual word in the phrase/syntax sequence are moved into the scanner and then sent into the transformer for scrutiny. To deal with cases of ambiguity, lateral inversion, and unnamed grammatical entities, the transformer shall be trained through part-of-speech tagging, in which

case, while recognizing several meanings of a given lemma as shall be drawn directly from C1 and D1. Contextual applications such as *l'ọ̀jī* or *lọ́jī* (passed head) in a more literary sense, but actually 'went over', or *gw'ọ̀jī* or gwọ̀jī (sit head) in a more literary sense, but actually 'sit on', upon conversion to English would then be represented in the transformer as *l'ọ̀jī* or *lọ́jī* => 'went over', while the syntactic form *gw'ọ̀jī* or gwọ̀jī => 'sit on'. In so doing, the transformer, even though it would receive a wide multiple meanings on the words being fed into it would be able to make prediction on the actual context the translation should appear, so that rather than interpreting a phrase like gwọ̀jī ėbījė as 'sit head iron', it becomes gwọ̀jī => {sit on} + {iron} => {sit on iron} => {sit on the iron}. You will notice the inking of the definite article 'the' with a tan tinted background. The reason is that upon parsing/annotation from Igala to English at A2 on figure VIII, the parser/annotator could not identify the word 'the', but since the syntactic form 'sit on chair' did not make a perfect sense in English, the word was

159

given a separate colour as a way of expressing explicit insertion that are not intrinsic in the word sequence of the source language, upon translation.

In probing for the unnamed grammatical entities, a similar annotation format - the BIO which is used for probing named entities in word prediction is used, although with a different purpose. The BIO annotation format – (inside, outside, beginning) was used in a parallel fashion to detect words and their meanings in the two language. Through the provision of the BIO annotated forms of syntax in the transformer, the transformer would be able to identify word sequence that comes closer to the meaning of another word sequence in the corresponding language, i.e. in respect to the phrase; *kā kū gbō* (say + that I + hear) 'say it let me hear', this phrase would be annotated using the BIO format as; say (B), it (O), let (O), me (O), hear (I). Through the application of a scanner with the capacity to detect both known and unknown entities during translation, it becomes possible for the translating machine to sense words that are not in the word sequence as translated which it thus, marks out using the tan colour pointed out above.

The two scanners attached to the transformers (the two tan-colour boxes attached to the transformer (the box with the vertical rectangle shape tied to C1, the two blue I boxes and the blue T to the left, and D1, the two orange colour I boxes and the yellow T box to the right)), are word detectors. Depending on the application though, either of these tan boxes send words in sequence as received from the transformer from either A2 or B2 through any of the two I on either side of figure VIII, following parsing and annotation. It then sense these words these words and their meanings through a crawling mechanism in more of a sense as the Google Search engine from either of the eight part-of-speech boxes on the two far flanks of the model for tagging and processing into meaning. The second tan boxes inside of the transformer; the two at the top, collects and returns unselected words and meanings from the transformer back to the part-of-speech boxes they evolved from in C1 and D1. When words are processed in the transformer, the translated equivalent are send via a pipe to the two T boxes on either side, for onward transmission to the screen as output.

5. Conclusion

The purpose of this paper therefore sprang from the need to create parsing trees and syntax annotation that could serve as bedrock of input materials that could be used for the development of a language corpus for the Igala, a needful resource that does not 'really' exist because previous efforts by Ayegba et al. (2017) are inadequate for want of extensive modeling

required while the paper by Joshua et al., (2020) does contain corpuses built on program interfaces, they are however not centered exclusively to corpuses and so they are not so comprehensive enough to serve the essence of that subject – corpus.

6. Acknowledgement

7. Ethical Statement

While conceiving a work of this nature, there are lots of issues to be dealt with such as the merit from which supportive datasets are extracted, the question as to whether available inputs that can be used to model a machine-based translator such as the tokens, lemmas, and the corpuses for an under-resourced language like the Igala where most of what exists as data needs extra verification before incorporation into other resource forms remains a biting question. Although while the accuracy of presentation and the approach adopted toward modeling by this author is not in doubt in a more literary sense, the fact that the was no provision for parsing contractive or agglutinative word forms by the Penn Treebank P.O.S tags led to the blending of these tags with a separate approach that was evolved by this very author – one that comes close to the tag sets and approach used by the Brown Corpus. Thus, in using the content of this paper, extra care should be taken by the consumers because of certain steps and approaches that might not be so accurate.

8. References

Abraham Sunday Unubi. (2017). "The use of

conjunctions in English and Igala: A contrastive analysis". *International Journal of Advanced Multidisciplinary Research*, 4(8), Pages 34-64.

Atadoga, F. T. (2011). 'Igala Morphology'. In Omachonu G. S (ed.) *Igala language Studies*. Saarbrucken: LAP Lambert Academic Publishing, Pages 76-102.

Attabor Theophilus Ocheja. (2019). "Syntactic Interference: A Study of Ígálá and English Noun Phrases in Malachai 1:6 and Mathew 2:1". *Journal of Literature, Languages and Linguistics*, 60, Pages 40-46.

Ayegba Sani Felix, Osuagwu O.E, and Njoku Dominic Okechukwu. (2014). "Machine Translation of Noun Phrases from English to Igala using the Rule-Based Approach". *West African Journal of Industrial and Academic Research*, 2(1): 18-28.

Ayegba Sani Felix, Abu Onoja, and Musa Ugbedeojo. (2017). "English – Igala Parallel Corpora for Natural Language Processing Applications". *International Journal of Computer Applications*, 171(9): 1-6.

Bonami Olivier, Godard Danièlle, and Marandin Jean-Marie. (1999). "Constituency and word order in French subject inversion". In Gosse Bouma, Erhard Hinrichs, Geert-Jan M. Kriujff, and Richard Oehrle (eds.) *Constraints and Resources in Natural Language Syntax and Semantics*. CSLI Publications, Stanford, USA, Pages 1-20.

Borsley D. Robert, Tallerman Maggie, and Willis David. (2007). *The Syntax of Welsh*. Cambridge University Press, Cambridge, UK, Pages 1-102.

Brunato Dominique, and Dell-Orletta Felice. (2017). "On the order of words in Italian: a study on genre vs complexity". In *Proceedings of the Fourth International Conference on Dependency Linguistics*, Pisa, Italy, September 18-20, Pages 25-31.

Comer J. William. (2021). "Russian's Most Frequent Words and Implications for Vocabulary Instruction," *Russian Language Journal*, 71(1),

Emmanuel Edoja Achor, and Christiana Akor. (2015). "Exploring the Use of Igala Language in Teaching Statistics to Samples of Selected Primary Six Learners". In *Proceedings of ISTE International Conference on "Towards Effective Teaching and Meaningful Learning in Mathematics, Science and Technology"*, University of South Africa, Pages 55-66.

Fransen Theodorus. (2020). "Automatic Morphological Parsing of Old Irish Verbs Using Finite-State Transducers". *Leeds Working Papers*, 1: 15-28.

Frazier Lyn. (1980). "Parsing and Constraints on Word Order". University of Massachusetts Occasional Papers, 6(5): 177-198.

Gudivada N. Venkat, and Arbabifard Kamyar. (2018). "Computational Analysis and Understanding of Natural Languages: Principles, Methods and Applications". In Gudivada N. Venkat, and Rao, C.R. *Handbook of Statistic 38*. https://www.sciencedirect.com/topics/computer-science/translation-model#:~:text=The%20language%20model%20can%20be,using%20the%20notion%20of%20alignments.

Hornby S. Albert. (2015). Oxford Advanced Learner's Dictionary of English, (Ninth Edition). Oxford University Press, Oxford, UK, Page 1210.

Ikani, F. E. (2011). 'Sense and Meaning Relations in Igala'. Omachonu G. S (ed.) Igala Language Studies. Saarbrucken: LAP Lambert Academic Publishing, Pages152-176.

Ileana Paul, and Postdam Eric. (2024). "Malagasy Framing Demonstratives and the Syntax of Doubling". *Glossa: A Journal of General Linguistics*, 9: 1-37.

Ilori, Folorunso. (2015). "Prepositions in Igala". *Ihafa: A Journal of African Studies*, 7(1): 139-160.

Joshua Attah, Ayegba Sani Felix, and Orah Richard Ojochegbe. (2020). "Example Based Machine Translation of English to Igala Language". *International Journal of Trend in Research and Development*, 7(1): Pages 32-35.

Jurafsky Daniel, and Martin H. James. (2009). Speech and Language Processing. https://en.wikipedia.org/wiki/Transfer-based_machine_translation#cite_note-slp-1

Kalin Laura. (2014). "The syntax of OVS word order in Hixkaryana". *Natural Language and Linguistic Theory*, 32: 1089-1104.

Mahmud Mohammed Momoh. (2023). "Vowels and the Igala Language Resources". In Proceedings of the Fourth workshop on Resources for African Indigenous Languages, pages 106-114.

Marcus Mitchell, Beatrice Santorini, and Marcinkiewicz Mary Ann. (1993). "Building a Large Annotated Corpus of English: The Penn Treebank". *Association for Computational Linguistics*, 19(2), pages 113-330.

Marcus Mitchell, Kim Grace, and Marcinkiewicz Mary Ann., et al. (1994). "The Penn Treebank: Annotating Predicate Argument Structure". https://aclanthology.org/H94-1020.pdf

McCone K. (1997). *The Early Irish verb. Revised edition with index verborum.* An Sagart, Maynooth, Ireland, Page 17.

Million Meshesha, and Yitayew Solomon. (2018). "English-Afaan Oromo Statistical Machine Translation". *International Journal of Computational Linguistic*, 9(1): 26-31.

Minhui Choi, and Emily Schmidt. (2015). "Postpositions and Word Order Variation in Korean". Linguistic Portfolios, 4(10): 108-115.

Mun Seongmin, and Desagulier Guillaume. (2022). "How do Transformer-Architecture Models Address Polysemy of Korean Adverbial

Postpositions?". In Proceedings of Deep Learning Inside Out (DEELIO 2022): The Third Workshop on Knowledge Extraction and Integration for Deep Learning Architectures, Dublin, Ireland, Pages 11-21.

Namboodiripad Savithry, Kim Dayoung, and Kim Gyeongnam. (2017). "English-dominant Korean-speakers show reduced flexibility in constituent order". *Chicago Linguistic Society*, 53:

Philpot, W.T.A. (1935). "Notes on the Igala Language". *Bulletin of the School of Oriental Studies*, 7(4): Pages 897-912.

Santorini Beatrice. (1990). *Part of Speech Tagging Guidelines for the Penn Treebank Project.*
https://www.cis.upenn.edu/~bies/manuals/tagguide.pdf

Wang, L., Tu, Z., and Zhang, X., et al. (2017). "A Novel and Robust Approach for Pro-drop Language Translation". *Machine Translation*, 31: 65-87.