Improving Accuracy of Igbo Corpus Annotation Using Morphological Reconstruction and Transformation-Based Learning

Ikechukwu Onyenwe, Mark Hepple & Uchechukwu Chinedu

NLP Group, Computer Science Department, University of Sheffield, UK

July 4, 2016

DESCRIPTION

PROCEDURE

EVALUATION and RESULT

SUMMARY

QUESTIONS

Research Aim

Research Aim

- **Contextual Information.** Taggers use context (information of the preceding words/tags of focus w) in order to disambiguate w correctly (Jurafsky & Martin, 2014). Quality of part-of-speech (POS) annotated corpus is really important here.

Research Aim

- **Contextual Information.** Taggers use context (information of the preceding words/tags of focus w) in order to disambiguate w correctly (Jurafsky & Martin, 2014). Quality of part-of-speech (POS) annotated corpus is really important here.
- **Tagger's performance.** Errors in a tagged corpus presents a threat to creating effective taggers. They give rise to "false context" that stand in place of "true context", which could have provided a good evidence in training taggers.

Research Aim

- **Contextual Information.** Taggers use context (information of the preceding words/tags of focus w) in order to disambiguate w correctly (Jurafsky & Martin, 2014). Quality of part-of-speech (POS) annotated corpus is really important here.
- **Tagger's performance.** Errors in a tagged corpus presents a threat to creating effective taggers. They give rise to "false context" that stand in place of "true context", which could have provided a good evidence in training taggers.
- **Error correction.** To develop efficient methods that will automatically detect likely errors in tagged corpora and possibly suggests plausible tags for correction which can be investigated by humans. The extensive labour of a human annotator going through the entire tagged texts methodically to find and correct errors will be greatly reduce.

About Igbo Language

About Igbo Language

- **Region/Speakers.** Igbo is one of the major languages spoken in eastern Nigeria by about 32 million native speakers (wikipedia).

About Igbo Language

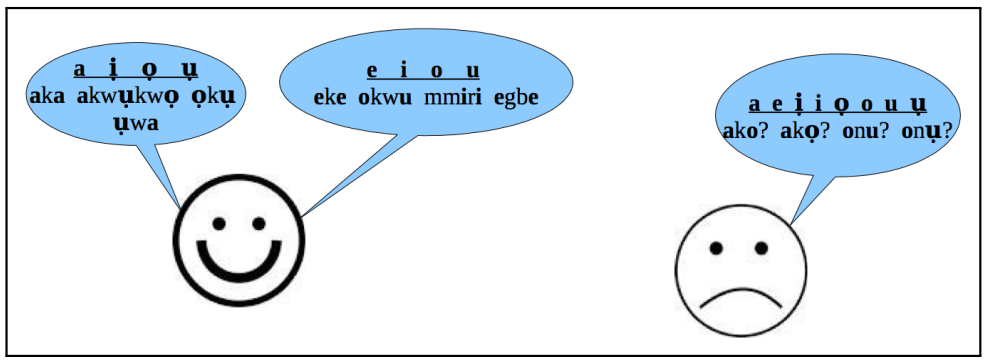
- **Region/Speakers.** Igbo is one of the major languages spoken in eastern Nigeria by about 32 million native speakers (wikipedia).
- **Classification.** It has been classified as a Benue-Congo language of the Kwa sub-group of the Niger-Congo family (igboguide.org).

About Igbo Language

- **Region/Speakers.** Igbo is one of the major languages spoken in eastern Nigeria by about 32 million native speakers (wikipedia).
- **Classification.** It has been classified as a Benue-Congo language of the Kwa sub-group of the Niger-Congo family (igboguide.org).
- **Orthography.** It adopts the Ọnwụ Committee orthography (onwu, 1961) and has 28 consonants and 8 vowels. Nine of the consonants are digraphs and 8 vowels divided in 2 harmony groups. The majority of the words of the language select their vowels from the same harmony group.

About Igbo Language

- **Region/Speakers.** Igbo is one of the major languages spoken in eastern Nigeria by about 32 million native speakers (wikipedia).
- **Classification.** It has been classified as a Benue-Congo language of the Kwa sub-group of the Niger-Congo family (igboguide.org).
- **Orthography.** It adopts the Ọnwụ Committee orthography (onwu, 1961) and has 28 consonants and 8 vowels. Nine of the consonants are digraphs and 8 vowels divided in 2 harmony groups. The majority of the words of the language select their vowels from the same harmony group.
- ...



Igbo Morphology

Igbo Morphology

- **Igbo is morphologically-rich language** A single stem in Igbo can produce as many possible word-forms as possible using affixes of varying lengths from 1 to 5 (**'a', 'ra', 'ghị', 'rịrị', 'gharị'**), which only extends the original meaning of the words.

Igbo Morphology

- **Igbo is morphologically-rich language** A single stem in Igbo can produce as many possible word-forms as possible using affixes of varying lengths from 1 to 5 (**'a', 'ra', 'ghị', 'rịrị', 'gharị'**), which only extends the original meaning of the words.
- For example, Illustrating word formation in Igbo using morphology

Word-form	Stem and Affixes	Meaning
ri	ri	eat
iri	i+ri	to eat
ga-eri	ga-+e+ri	will eat (AV hyphenated to participle)
ga-ericha	ga-+e+ri+cha	will eat completely
ga-erichariri	ga-+e+ri+cha+rịrị	will must eat completely
ga-erikwa	ga-+e+ri+kwa	will eat also
richariri	ri+cha+rịrị	must eat completely
richakwa	ri+cha+kwa	eat completely also
richara	ri+cha+ra	ate completely
richakwara	ri+cha+kwa+ra	ate completely also

NLP Resources Available in Igbo

Igbo Tagset and about 1 million sized corpus (made of religious and modern Igbo texts genres). About 260k of 1m is POS-tagged. Methods reported in (Onyenwe et al., 2014, 2015) @ LAW VIII COLING14 and Joint Workshop on Language Technology RANLP 2015.

Error Correction Methods and Morphological Analysis

Error Correction Methods and Morphological Analysis

- There have been previous works done in correcting errors found automatically in a tagged corpus. Instead of going through tagged corpora methodically by human annotators to find and correct errors, an efficient means can be developed that uses the human annotator expert in its process loop to correct errors found. **Examples**: Brill & Marcus, 1992; Taljard et al., 2008; Heid et al., 2006; Loftsson, 2009; Helgadóttir et al., 2012; Leech et al., 1983; etc applied automatic methods that find wrongly assigned tags in tagged corpora, involving humans to validates the positions in tagged corpora the automatic methods have identified to be erroneous.

Error Correction Methods and Morphological Analysis

- There have been previous works done in correcting errors found automatically in a tagged corpus. Instead of going through tagged corpora methodically by human annotators to find and correct errors, an efficient means can be developed that uses the human annotator expert in its process loop to correct errors found. **Examples**: Brill & Marcus, 1992; Taljard et al., 2008; Heid et al., 2006; Loftsson, 2009; Helgadóttir et al., 2012; Leech et al., 1983; etc applied automatic methods that find wrongly assigned tags in tagged corpora, involving humans to validates the positions in tagged corpora the automatic methods have identified to be erroneous.
- Morphological analysis has been usefully exploited elsewhere in natural language processing. **Examples**: Thede et al., 1997, investigated whether morphological information could assist in handling unknown words in the context of syntactic parsing. Milne 1986 used morphological reconstruction to resolve ambiguity during parsing. Light, 1996 used various knowledge sources to determine word meanings, including morphological cues.

Current State of IgbTC

Current State of IgbTC

- After corpus was first annotated, tagset was revised, and IAA exercise carried out. Materials from IAA exercise used with ML (TBL) to:

Current State of IgbTC

- After corpus was first annotated, tagset was revised, and IAA exercise carried out. Materials from IAA exercise used with ML (TBL) to:
  - propagate revised tags to corpus

Current State of IgbTC

- After corpus was first annotated, tagset was revised, and IAA exercise carried out. Materials from IAA exercise used with ML (TBL) to:
  - propagate revised tags to corpus
  - identify likely cases where errors made (e.g. where IAA annotators disagree), for human expert inspection.

Current State of IgbTC

- After corpus was first annotated, tagset was revised, and IAA exercise carried out. Materials from IAA exercise used with ML (TBL) to:
  - propagate revised tags to corpus
  - identify likely cases where errors made (e.g. where IAA annotators disagree), for human expert inspection.
- Used committee of taggers (COT) (Loftsson, '09 & Helgadóttir etal., '12) where two or more taggers agreed on a tag but different from what is in the gold standard (a good candidate for inspection).

Current State of IgbTC

Total statistics outcomes of the improvement methods.

Name	Location Flagged	Accepted Judgement	No-Change Required	Manual Change	Effective Change	% Error Removed
IAA+TBL 1	25490	16612	5569	3309	19921	7.550
IAA+TBL 2	26155	3605	20471	2079	5684	2.154
COT	11810	6549	4165	1096	7645	2.897
Total	63455	26766	30205	6484	33250	12.601

Current State of IgbTC

Total statistics outcomes of the improvement methods.

Name	Location Flagged	Accepted Judgement	No-Change Required	Manual Change	Effective Change	% Error Removed
IAA+TBL 1	25490	16612	5569	3309	19921	7.550
IAA+TBL 2	26155	3605	20471	2079	5684	2.154
COT	11810	6549	4165	1096	7645	2.897
Total	63455	26766	30205	6484	33250	12.601

- IAA+TBL 1: 78% of 25490 were effectively changed, i.e., 7.550% errors were eliminated from IgbTC. IAA+TBL 2: The rate of corrections ($\approx$22%) here is substantially lower than 1, but 2.154% errors were eliminated from IgbTC justifies its effectiveness. Same for COT.

Current State of IgbTC

Total statistics outcomes of the improvement methods.

Name	Location Flagged	Accepted Judgement	No-Change Required	Manual Change	Effective Change	% Error Removed
IAA+TBL 1	25490	16612	5569	3309	19921	7.550
IAA+TBL 2	26155	3605	20471	2079	5684	2.154
COT	11810	6549	4165	1096	7645	2.897
Total	63455	26766	30205	6484	33250	12.601

- IAA+TBL 1: 78% of 25490 were effectively changed, i.e., 7.550% errors were eliminated from IgbTC. IAA+TBL 2: The rate of corrections (≈22%) here is substantially lower than 1, but 2.154% errors were eliminated from IgbTC justifies its effectiveness. Same for COT.
- Overall, 24.05% of IgbTC with 12.601% effective change made and accuracy increased from 88% (initial state) to 96% (current state) obtained by training and testing FnTBL tagger on IgbTC sets on 10-fold cross validation over the corpus size.

Current State of IgbTC

Total statistics outcomes of the improvement methods.

Name	Location Flagged	Accepted Judgement	No-Change Required	Manual Change	Effective Change	% Error Removed
IAA+TBL 1	25490	16612	5569	3309	19921	7.550
IAA+TBL 2	26155	3605	20471	2079	5684	2.154
COT	11810	6549	4165	1096	7645	2.897
Total	63455	26766	30205	6484	33250	12.601

- IAA+TBL 1: 78% of 25490 were effectively changed, i.e., 7.550% errors were eliminated from IgbTC. IAA+TBL 2: The rate of corrections (≈22%) here is substantially lower than 1, but 2.154% errors were eliminated from IgbTC justifies its effectiveness. Same for COT.
- Overall, 24.05% of IgbTC with 12.601% effective change made and accuracy increased from 88% (initial state) to 96% (current state) obtained by training and testing FnTBL tagger on IgbTC sets on 10-fold cross validation over the corpus size.
- About 260k of main corpus has been tagged and improved (Onyenwe et al., 2014, 2015).

Experimental Data and Tools

Aim: Igbo tagset is designed to have two parts separated by hyphen(α_XS), α is non morphologically-inflected part and α_XS, where _XS indicates morphologically-inflected. To developed an automatic method that find errors where the assignment of tags violates the status of words that are morphologically-inflected in IgbTC. Tools used

Experimental Data and Tools

Aim: Igbo tagset is designed to have two parts separated by hyphen(α_XS), α is non morphologically-inflected part and α_XS, where _XS indicates morphologically-inflected. To developed an automatic method that find errors where the assignment of tags violates the status of words that are morphologically-inflected in IgbTC. Tools used

- Transformation-based learning in the fast lane (FnTBL) by Ngai & Florian (2001), a reimplementation of Brill's TBL because of linguistic pattern detection and Stanford Log-linear Tagger (SLLT) (Toutanova et al., 2003). SLLT was applied on the outcome of this experiment because it has robustness in word feature extraction.

Experimental Data and Tools

Aim: Igbo tagset is designed to have two parts separated by hyphen(α_XS), α is non morphologically-inflected part and α_XS, where _XS indicates morphologically-inflected. To developed an automatic method that find errors where the assignment of tags violates the status of words that are morphologically-inflected in IgbTC. Tools used

- Transformation-based learning in the fast lane (FnTBL) by Ngai & Florian (2001), a reimplementation of Brill's TBL because of linguistic pattern detection and Stanford Log-linear Tagger (SLLT) (Toutanova et al., 2003). SLLT was applied on the outcome of this experiment because it has robustness in word feature extraction.
- Morphological Reconstruction (MR), a linguistically-informed segmentation of words into roots and affixes, the knowledge of the roots and associated affixes are used to process words (especially unknown words) (Thede & Harper, 1997).

Experimental Data and Tools

Aim: Igbo tagset is designed to have two parts separated by hyphen(α_XS), α is non morphologically-inflected part and α_XS, where _XS indicates morphologically-inflected. To developed an automatic method that find errors where the assignment of tags violates the status of words that are morphologically-inflected in IgbTC. Tools used

- Transformation-based learning in the fast lane (FnTBL) by Ngai & Florian (2001), a reimplementation of Brill's TBL because of linguistic pattern detection and Stanford Log-linear Tagger (SLLT) (Toutanova et al., 2003). SLLT was applied on the outcome of this experiment because it has robustness in word feature extraction.
- Morphological Reconstruction (MR), a linguistically-informed segmentation of words into roots and affixes, the knowledge of the roots and associated affixes are used to process words (especially unknown words) (Thede & Harper, 1997).
- The current state of IgbTC sized 263856 tokens.

Methods

Methods

- Develop morphological segmentation (MS) Using MR, performs segmentation of morphologically-inflected words in IgbTC and classify the segmentation outputs. We only focused on verbal class at this phase since it constitute the majority of inflected class and to avoid performing full computational morphology in Igbo (it wll take ages).

Methods

- Develop morphological segmentation (MS) Using MR, performs segmentation of morphologically-inflected words in IgbTC and classify the segmentation outputs. We only focused on verbal class at this phase since it constitute the majority of inflected class and to avoid performing full computational morphology in Igbo (it wll take ages).
- Before MS takes place, words that are not in verbal words that are morphologically-inflected are sieved out. Discuss this later ...

Methods

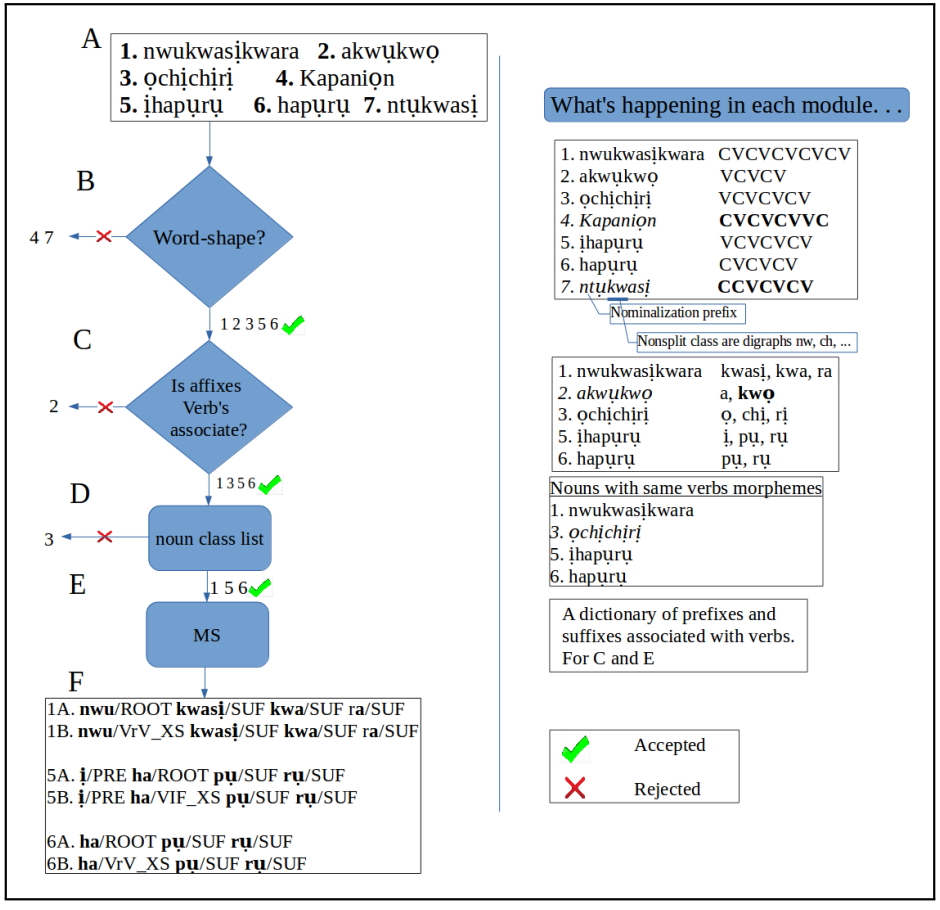
- Develop morphological segmentation (MS) Using MR, performs segmentation of morphologically-inflected words in IgbTC and classify the segmentation outputs. We only focused on verbal class at this phase since it constitute the majority of inflected class and to avoid performing full computational morphology in Igbo (it wll take ages).
- Before MS takes place, words that are not in verbal words that are morphologically-inflected are sieved out. Discuss this later ...
- The outputs of MS serves as FnTBL's initial and truth states. IgbTC was sets on 10-fold cross validation over the corpus size. 90% used for training on verbal words in IgbTC that are morphologically-inflected. The trained FnTBL are used on the remainder.

Methods

- Develop morphological segmentation (MS) Using MR, performs segmentation of morphologically-inflected words in IgbTC and classify the segmentation outputs. We only focused on verbal class at this phase since it constitute the majority of inflected class and to avoid performing full computational morphology in Igbo (it wll take ages).
- Before MS takes place, words that are not in verbal words that are morphologically-inflected are sieved out. Discuss this later ...
- The outputs of MS serves as FnTBL's initial and truth states. IgbTC was sets on 10-fold cross validation over the corpus size. 90% used for training on verbal words in IgbTC that are morphologically-inflected. The trained FnTBL are used on the remainder.
- The plan: to use the morphological clues to predict the correct tags for the morphologically-inflected words that are verbs, where predicted tags disagree with tags in IgbTC will flagged for inspection.

Methods

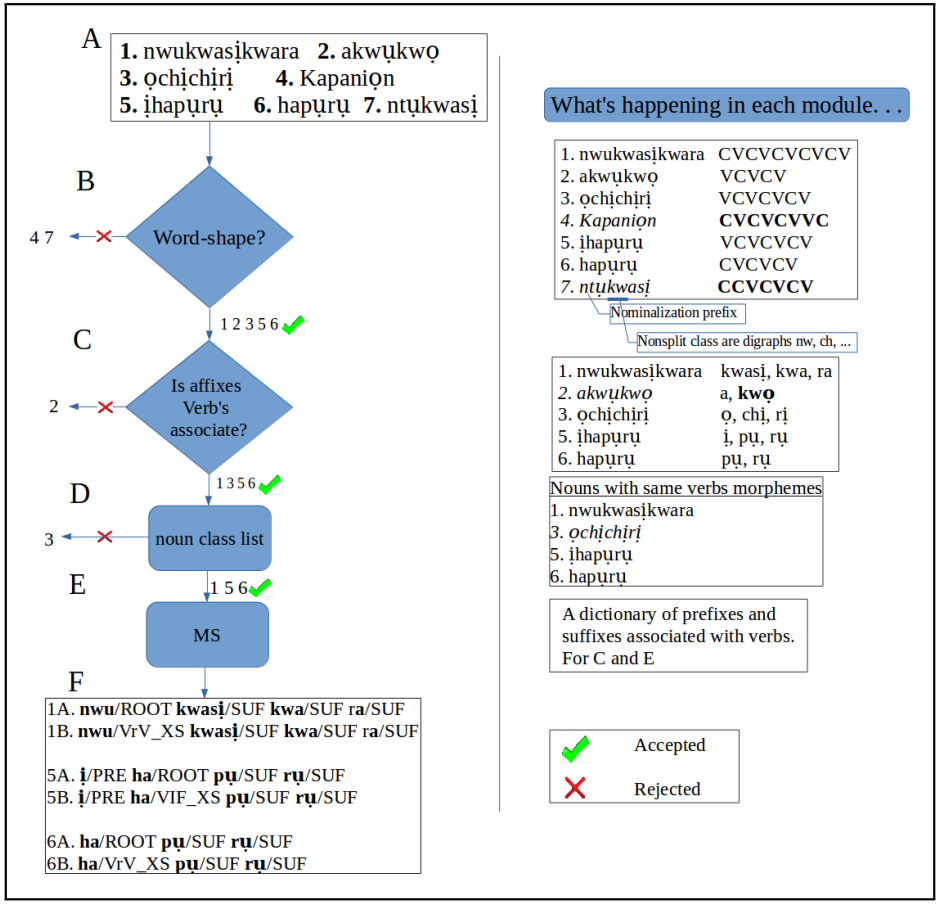
Flow diagram of operation: identify morphologically-inflected verbs, perform morphological reconstruction and classification.



Methods

Flow diagram of operation: identify morphologically-inflected verbs, perform morphological reconstruction and classification.

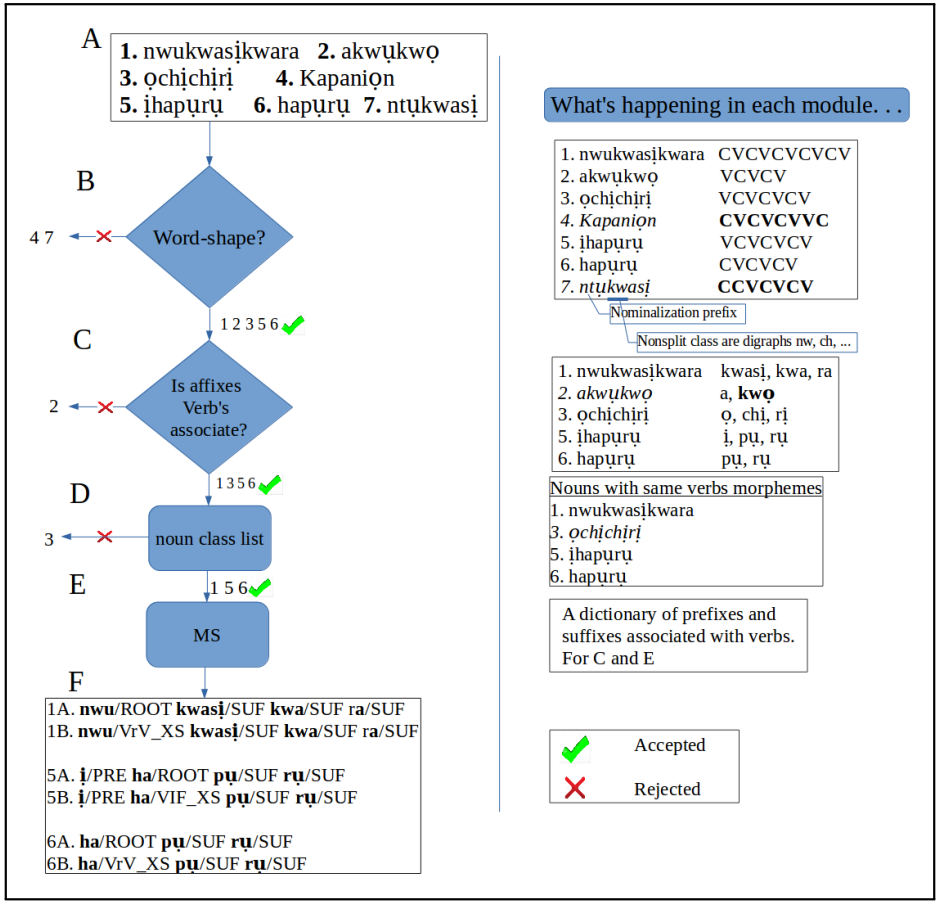
- **A:** Inputs from IgbTC.



Methods

Flow diagram of operation: identify morphologically-inflected verbs, perform morphological reconstruction and classification.

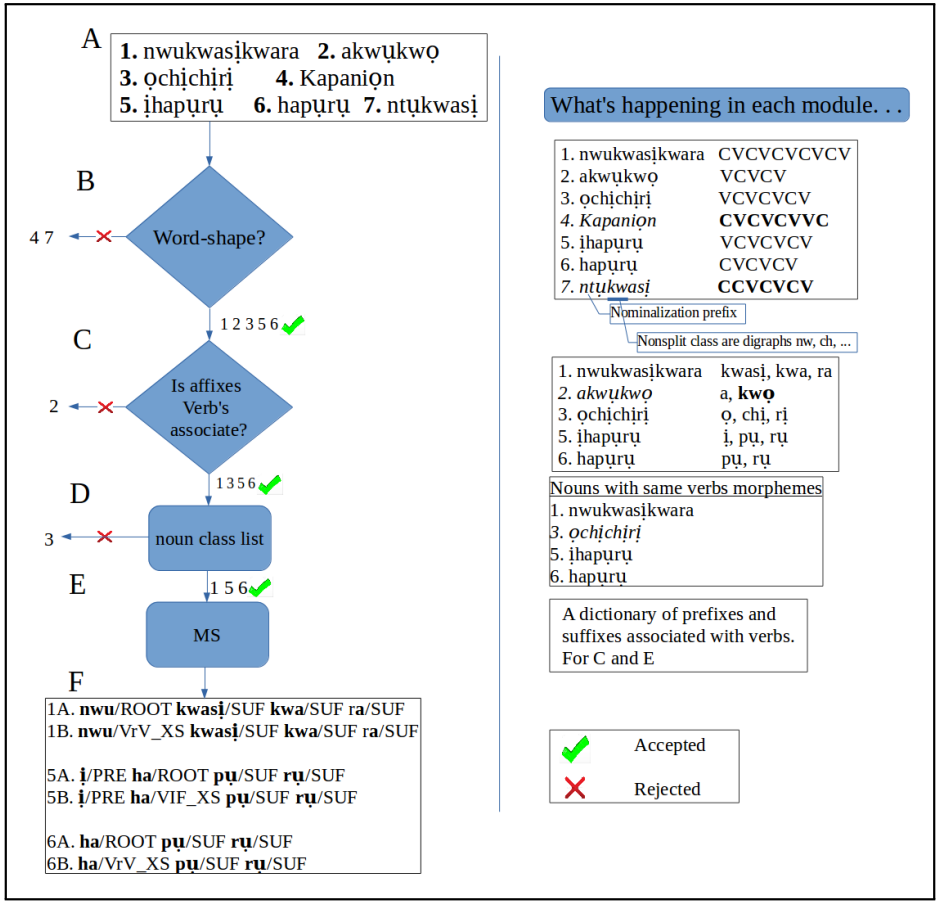
- **A:** Inputs from IgbTC.
- **B:** Consonant C and Vowel V.



Methods

Flow diagram of operation: identify morphologically-inflected verbs, perform morphological reconstruction and classification.

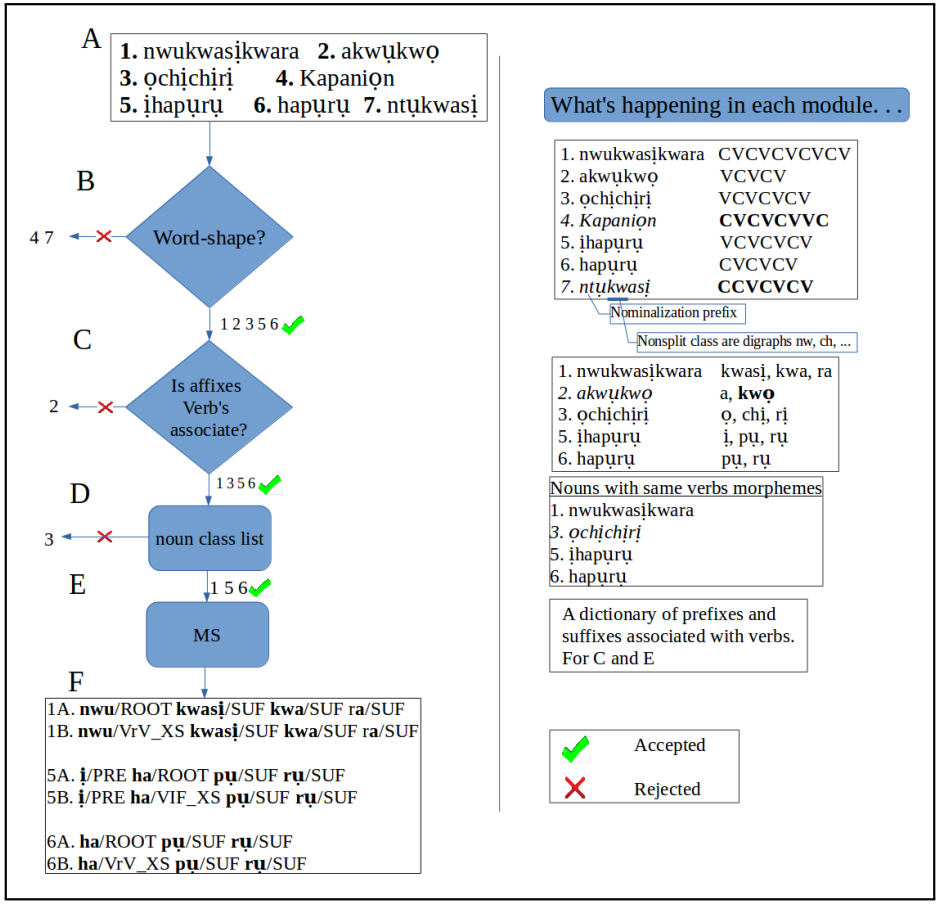
- **A:** Inputs from IgbTC.
- **B:** Consonant C and Vowel V.
- **C:** Valid verbs' Affixes.



Methods

Flow diagram of operation: identify morphologically-inflected verbs, perform morphological reconstruction and classification.

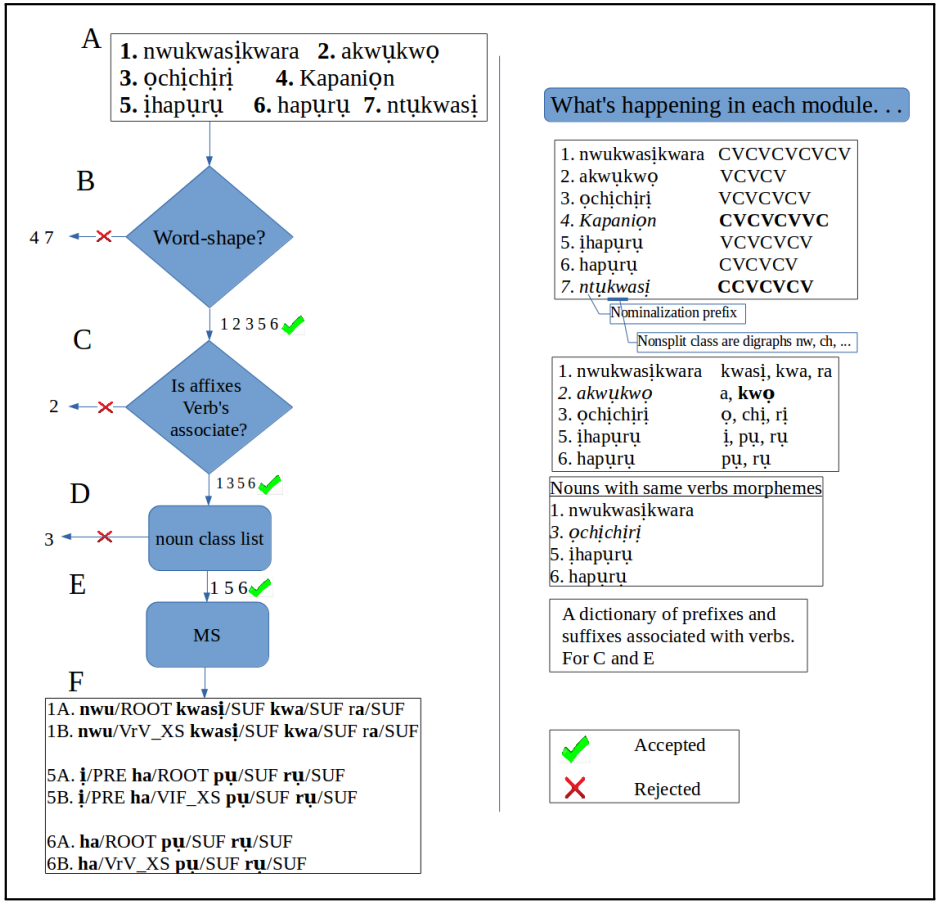
- **A:** Inputs from IgbTC.
- **B:** Consonant C and Vowel V.
- **C:** Valid verbs' Affixes.
- **D:** Removes remaining non verbs.



Methods

Flow diagram of operation: identify morphologically-inflected verbs, perform morphological reconstruction and classification.

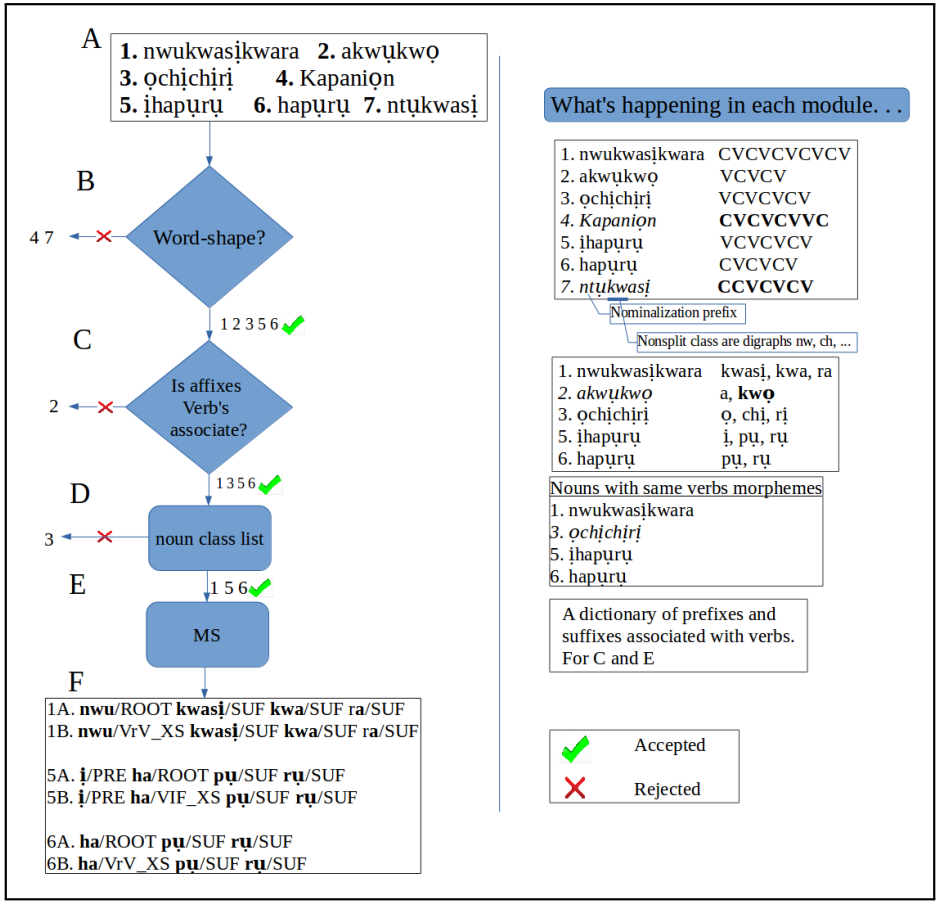
- **A:** Inputs from IgbTC.
- **B:** Consonant C and Vowel V.
- **C:** Valid verbs' Affixes.
- **D:** Removes remaining non verbs.
- **E:** MR and Classification.



Methods

Flow diagram of operation: identify morphologically-inflected verbs, perform morphological reconstruction and classification.

- **A:** Inputs from IgbTC.
- **B:** Consonant C and Vowel V.
- **C:** Valid verbs' Affixes.
- **D:** Removes remaining non verbs.
- **E:** MR and Classification.
- **F:** FnTBL's initial and truth states.



FnTBL error correction process using morphological cues

Some interesting sample rules:

FnTBL error correction process using morphological cues

Some interesting sample rules:

- pos_0=ROOT pos:[1,3]=SUF => pos=VrV, generic rule to transform all ROOT to "VrV" (past tense verbs) if there is any suffix.

FnTBL error correction process using morphological cues

Some interesting sample rules:

- pos_0=ROOT pos:[1,3]=SUF => pos=VrV, generic rule to transform all ROOT to "VrV" (past tense verbs) if there is any suffix.
- pos_0=VrV word_1=kwasị word_2=kwa word_3=ra => pos=VrV_XS, transform VrV to VrV_XS if suffixes after stem is "kwasị", "kwa" and "ra". E.g. *nwukwasịkwara*, *bịakwasịkwara*, *dakwasịkwara*, ... "ra" marked them past tense verbs, making it different from simple inflected verbs.

FnTBL error correction process using morphological cues

Some interesting sample rules:

- pos_0=ROOT pos:[1,3]=SUF => pos=VrV, generic rule to transform all ROOT to "VrV" (past tense verbs) if there is any suffix.
- pos_0=VrV word_1=kwasị word_2=kwa word_3=ra => pos=VrV_XS, transform VrV to VrV_XS if suffixes after stem is "kwasị", "kwa" and "ra". E.g. *nwukwasịkwara*, *bịakwasịkwara*, *dakwasịkwara*, ... "ra" marked them past tense verbs, making it different from simple inflected verbs.
- pos_0=VrV_XS word_-1=ị => pos=VIF_XS, transform VrV_XS to VIF_XS if prefix is ị. ị marks infinitive verbs in Igbo. Compare *hapụrụ* "left for" and *ịhapụrụ* "to leave for ".

FnTBL error correction process using morphological cues

Some interesting sample rules:

- pos_0=ROOT pos:[1,3]=SUF => pos=VrV, generic rule to transform all ROOT to "VrV" (past tense verbs) if there is any suffix.
- pos_0=VrV word_1=kwasị word_2=kwa word_3=ra => pos=VrV_XS, transform VrV to VrV_XS if suffixes after stem is "kwasị", "kwa" and "ra". E.g. *nwukwasịkwara*, *bịakwasịkwara*, *dakwasịkwara*, ... "ra" marked them past tense verbs, making it different from simple inflected verbs.
- pos_0=VrV_XS word_-1=ị => pos=VIF_XS, transform VrV_XS to VIF_XS if prefix is ị. ị marks infinitive verbs in Igbo. Compare *hapụrụ* "left for" and *ịhapụrụ* "to leave for ".
- pos_0=VSI_XS word:[1,2]=wo => pos=VPERF, transform VSI_XS to VPERF if part of speech is VSI_XS and suffix within the range of 1-2 after stem is "wo". "wo","go" and "la" are perfect tense marker. This rule could also be pos_0=VSI_XS word:[1,2]=ZZZ => pos=VPERF, where ZZZ could be any of "wo","go" and "la". "ọ gbawo egwu" "he has danced".

FnTBL error correction process

Igbo tagset has α_*XS* form, while if α, is non morphologically-inflected part and if α_*XS*, _XS indicates morphologically-inflected. So, if α in both tags of a flagged position in IgbTC are same and there is any SUF, FnTBL predicted tag will be accepted. Others manually corrected. For quality assurance, all flagged positions were inspected by a human annotator expert. Sample output below

FnTBL error correction process

Igbo tagset has α_*XS* form, while if α, is non morphologically-inflected part and if α_*XS*, _XS indicates morphologically-inflected. So, if α in both tags of a flagged position in IgbTC are same and there is any SUF, FnTBL predicted tag will be accepted. Others manually corrected. For quality assurance, all flagged positions were inspected by a human annotator expert. Sample output below

IgbTC Before Error Correction	IgbTC After Error Correction
nwukwasịkwara/VrV	nwukwasịkwara/VrV_XS
pụkwaghị/VrV_XS	pụkwaghị/VSI_XS
bụrụkwa/VrV_XS	bụrụkwa/VSI_XS
laara/VrV	laara/VrV_XS
waara/VrV	waara/VrV_XS
zooro/VrV	zooro/VrV_XS
zukwaara/VrV	zukwaara/VrV_XS
kwughachikwa/VCO	kwughachikwa/VSI_XS
kwuluwo/VSI_XS	kwuluwo/VPERF
ịhapụrụ/VrV_XS	ịhapụrụ/VIF_XS

For evaluation, SLLT was applied on IgbTC after error correction sets on 10-fold cross validation.

IgbTC State Before			IgbTC State After		
Overall Sc	Unknown Sc	MIU Sc	Overall Sc	Unknown Sc	MIU Sc
98.05%	77.77%	58.01%	98.11%	83.43%	86.81%

MIU is morphologically-inflected unknown words.

For evaluation, SLLT was applied on IgbTC after error correction sets on 10-fold cross validation.

IgbTC State Before			IgbTC State After		
Overall Sc	Unknown Sc	MIU Sc	Overall Sc	Unknown Sc	MIU Sc
98.05%	77.77%	58.01%	98.11%	83.43%	86.81%

MIU is morphologically-inflected unknown words.

- SLLT accuracy scores on IgbTC generally increased. The effect is very prominent in the accuracy of the unknown words (rare words), especially the morphologically-inflected class (0.06% for overall, 5.66% for unknown words and 28.8% for morphologically-inflected words that are unknown).

For evaluation, SLLT was applied on IgbTC after error correction sets on 10-fold cross validation.

IgbTC State Before			IgbTC State After		
Overall Sc	Unknown Sc	MIU Sc	Overall Sc	Unknown Sc	MIU Sc
98.05%	77.77%	58.01%	98.11%	83.43%	86.81%

MIU is morphologically-inflected unknown words.

- SLLT accuracy scores on IgbTC generally increased. The effect is very prominent in the accuracy of the unknown words (rare words), especially the morphologically-inflected class (0.06% for overall, 5.66% for unknown words and 28.8% for morphologically-inflected words that are unknown).
- Following this outcome, we can hypothesize that SLLT performance on "IgbTC State Before" is due to

For evaluation, SLLT was applied on IgbTC after error correction sets on 10-fold cross validation.

IgbTC State Before			IgbTC State After		
Overall Sc	Unknown Sc	MIU Sc	Overall Sc	Unknown Sc	MIU Sc
98.05%	77.77%	58.01%	98.11%	83.43%	86.81%

MIU is morphologically-inflected unknown words.

- SLLT accuracy scores on IgbTC generally increased. The effect is very prominent in the accuracy of the unknown words (rare words), especially the morphologically-inflected class (0.06% for overall, 5.66% for unknown words and 28.8% for morphologically-inflected words that are unknown).
- Following this outcome, we can hypothesize that SLLT performance on "IgbTC State Before" is due to
  - likely using "false context" learnt from the errors in the corpus in its prediction.

For evaluation, SLLT was applied on IgbTC after error correction sets on 10-fold cross validation.

IgbTC State Before			IgbTC State After		
Overall Sc	Unknown Sc	MIU Sc	Overall Sc	Unknown Sc	MIU Sc
98.05%	77.77%	58.01%	98.11%	83.43%	86.81%

MIU is morphologically-inflected unknown words.

- SLLT accuracy scores on IgbTC generally increased. The effect is very prominent in the accuracy of the unknown words (rare words), especially the morphologically-inflected class (0.06% for overall, 5.66% for unknown words and 28.8% for morphologically-inflected words that are unknown).
- Following this outcome, we can hypothesize that SLLT performance on "IgbTC State Before" is due to
  - likely using "false context" learnt from the errors in the corpus in its prediction.
  - Some SLLT's classifications are likely to be correct but were penalized because of these errors.

For evaluation, SLLT was applied on IgbTC after error correction sets on 10-fold cross validation.

IgbTC State Before			IgbTC State After		
Overall Sc	Unknown Sc	MIU Sc	Overall Sc	Unknown Sc	MIU Sc
98.05%	77.77%	58.01%	98.11%	83.43%	86.81%

MIU is morphologically-inflected unknown words.

- SLLT accuracy scores on IgbTC generally increased. The effect is very prominent in the accuracy of the unknown words (rare words), especially the morphologically-inflected class (0.06% for overall, 5.66% for unknown words and 28.8% for morphologically-inflected words that are unknown).
- Following this outcome, we can hypothesize that SLLT performance on "IgbTC State Before" is due to
  - likely using "false context" learnt from the errors in the corpus in its prediction.
  - Some SLLT's classifications are likely to be correct but were penalized because of these errors.
- The corrections are mainly rare words with less frequency, rare occurrence caused by morphology. *The more suffixes ... less frequency (rare).

Summary of this Research

This is how we used stems and associated affixes to transform wrong tags assigned to morphologically-inflected words to their true tags in tagged Igbo corpus (IgbTC). Morphological reconstruction was used to represent these words in IgbTC in machine learnable pattern that FnTBL exploited to identify wrongly tagged ones and suggest plausible tags for correction. Human annotator expert inspected all the affected positions on IgbTC for quality assurance.

Questions and Answers

THANKS FOR LISTENING. Questions?