Springer Link

Addressing the Challenges of Igbo Computational Morphological Studies Using Frequent Pattern-Based Induction

- Authors
- Authors and affiliations
- Olamma U. Iheanetu
- Obododimma Oha

Abstract

Computational studies of Igbo language are constrained by non-availability of large electronic corpora of Igbo text, a prerequisite for data-driven morphological induction. Existing unsupervised models, which are frequent-segment based, do not sufficiently address non-concatenative morphology and cascaded affixation prevalent in Igbo morphology, as well achieving affix labelling. This study devised a data-driven model that could induce non-concatenative aspects of Igbo morphology, cascaded affixation and affix labelling using frequent pattern-based induction. Ten-fold Cross Validation (TCV) test was used to validate the propositions using percentages. An average accuracy measure of 88% was returned for the developed model. Ten purposively selected Igbo first speakers also evaluated samples of 100 model-analysed words each and the mean accuracy score of 82% was recorded. We conclude that morphology induction can be realized with a modestly sized corpus, demonstrating that electronic corpora scarcity does not constrain computational morphology studies as it would other higher levels of linguistic analysis.

Keywords

This is a preview of subscription content, log in to check access.

Notes

Acknowledgements

References

1. 1.

O. Awobuliyi, *Eko Iseda-Oro Yoruba* (Montem Paper Backs, Akure, Ondo state, 2008)Google Scholar

2. 2.

K.R. Beesley, L. Karttunen, *Finite State Morphology* (CSLI Publications, Stanford, United States of America, 2003)Google Scholar

3. 3.

R. Blench, *Atlas of Nigerian languages*, 3rd edn. (2012) Retrieved from 9 June 2015, www.rogerblench.info/Language/Africa/Nigeria/Atlas%20of%20Nigerian%20Languages-%20ed%20III.pdf

4. 4.

M. Creutz, Induction of the morphology of natural language: unsupervised morpheme segmentation with application to automatic speech recognition. Ph.D. Thesis, Computer and Information Science Department, Helsinki, University of Technology, Espoo (2007), xi+110 pp.Google Scholar

5. 5.

G. De Pauw, G. De Schryver, Improving the computational morphological analysis of a Swahili corpus for lexicographic purposes. Lexikos Afr. Assoc. Lexicogr. (AFRILEX)-reeks Series **18**, 303–318 (2008)Google Scholar

6. 6.

N. Emenanjo, The interfix: an aspect of universal morphology. J. West Afr. Lang. XII **1**(1982), 77–88 (1982)Google Scholar

7. 7.

N. Emenanjo, *Elements of Modern Igbo Grammar* (University Press Limited (UPL), Ibadan, 1987)Google Scholar

8. 8.

M.A. Fullwood, T.J. O'Donnell, Learning non-concatenative morphology, in *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, 8 August 2013, Sofia, Bulgaria, pp. 21–27Google Scholar

9. 9.

Gale Group Inc., Igbo. Junior Worldmark Encyclopedia of World Cultures (1999). Retrieved from 10 August 2010, Encyclopedia.com: http://www.encyclopedia.com/doc/1G2-3435900354.html

10. 10.

J. Goldsmith, Unsupervised learning of the morphology of a natural language. Mass. Inst. Technol. (MIT) Press J. **27**(2), 153–198

(2001). https://doi.org/10.1162/089120101750300490MathSciNetCrossRefGoogle Scholar

11. 11.

J. Goldsmith, An algorithm for the unsupervised learning of morphology. Nat. Lang. Eng. **1**(1) (2005). Cambridge University Press. Retrieved from, http://hum.uchicago.edu/~jagoldsm/Papers/algorithm.pdf

12. 12.

H. Hammarström, Unsupervised learning of morphology and the languages of the world, Doctoral thesis, Department of Computer Science and Engineering, Chalmers University of Technology and University of Gothenburg, Sweden (2009), 284 pp.Google Scholar

13. 13.

H. Hammarström, L. Borin, Unsupervised learning of morphology. MIT Press J. **37**(2), 309–350 (2010)Google Scholar

14. 14.

Z. Harris, *Morpheme Boundaries within Words: Report on a Computer Test.* Transformations and Discourse Analysis Papers (1967), p. 73Google Scholar

15. 15.

O.U. Iheanetu, Data-driven model of Igbo morphology. Doctoral thesis, Africa Regional Centre for Information Science, (ARCIS), University of Ibadan, Nigeria (2015) 284 pp.Google Scholar

16. 16.

O.U. Iheanetu, O. Oha, Some salient issues in the unsupervised learning of Igbo morphology, in *World Congress on Engineering and Computer Science 2017*. Lecture Notes in Engineering and Computer Science, 25–27 October 2017, San Francisco, USA, pp. 389–393Google Scholar

17. 17.

P. Lambert, M. Costa-jussa, R.E. Banchas, Introduction, in *Workshop on Creating Cross-Language Resources for Disconnected Languages and Styles*, 27th May 2012. Istanbul, Turkey (2012)Google Scholar

18. 18.

J.J. McCarthy, A prosodic theory of nonconcatenative morphology. Linguist. Inq. **12**(3), 373–418 (1981)Google Scholar

19. 19.

P. MacClanahan, G. Busby, R. Haertel, K. Heal, D. Lonsdale, K. Seppi, E. Ringer, A probabilistic morphological analyser for Syriac, in *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (2010), pp. 810–820Google Scholar

20. 20.

B.M. Mba, *A Minimalist Theory and Application to Igbo* (Catholic Institute for Development Justice and Peace (CIDJAP) Press, Enugu, 2011)Google Scholar

21. 21.

B.M. Mba, Circumfixation: interface of morphology and syntax in Igbo derivational morphology. IOSR J. Humanit. Soc. Sci. (JHSS) **5**(6), 1–8 (2012)CrossRefGoogle Scholar

22. 22.

O.M. Ndimele, *A First Course on Morphology and Syntax* (Emhai Printing and Publishing Company, Port Harcourt, 1999)Google Scholar

23. 23.

B.I.N. Osuagwu, G.I. Nwaozuzu, G.A. Dike, V.N. Nwaogu, L.C. Okoro, *Fundamentals of linguistics* (Colon Concept Ltd, Owerri, 1997)Google Scholar

24. 24.

L.M. Paul, G.F. Simons, C.D. Fennig (eds.), *Ethnologue: Languages of the World, Eighteenth Edition* (SIL International, Dallas, Texas, 2015). Retrieved from 20 June 2015, http://www.ethnologue.com/language/ibo

25. 25.

A.K. Simpson, The origin and development of nonconcatenative morphology. Ph.D. Thesis. Graduate Division of the Department of California (2009), 194 pp.Google Scholar

26. 26.

University of California, Los Angeles (UCLA) Language Materials Project 2009. Igbo. UCLA Language Materials Project. Retrieved from 20 October 2010, http://www.lmp.ucla.edu/Profile.aspx?LangID=13&menu=004

27. 27.

V.N. Vapnik, An overview of statistical learning theory. IEEE Trans. Neural Netw. **10**(5), 988–999 (1999)CrossRefGoogle Scholar

Copyright information

About this paper

CrossMark

Buy eBook
EUR 160.49

Buy paper (PDF)

-