

# Improving morphotactic inference when roots aren't identified

Tara Wueger, University of Washington  
DELPH-IN Summit 2022



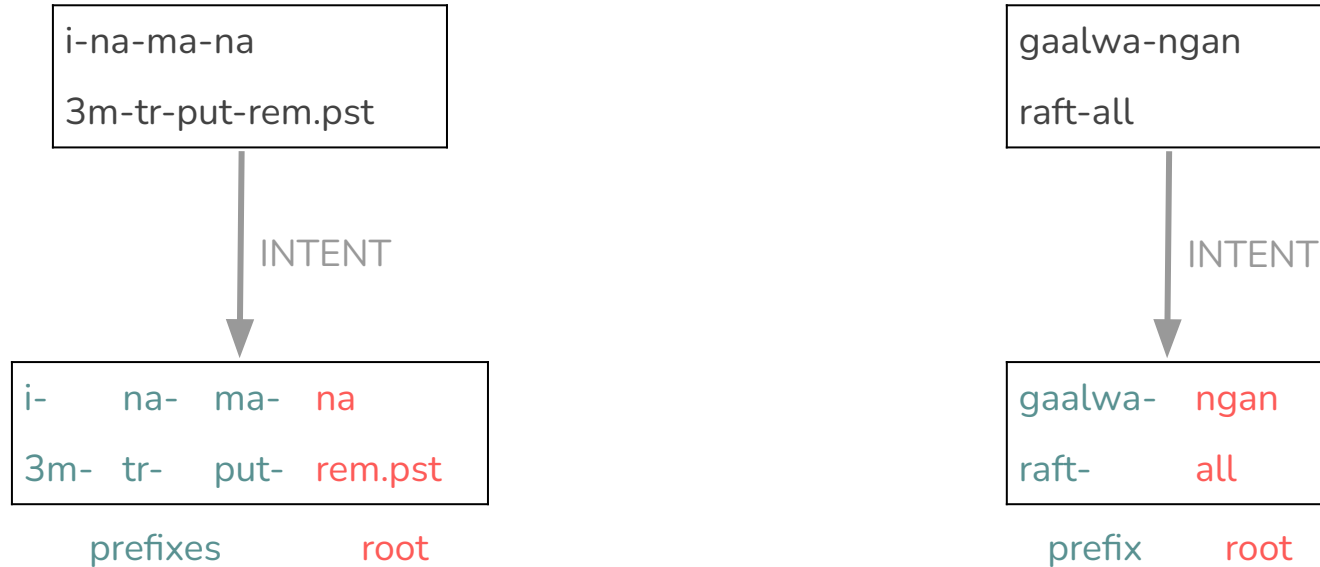


## Issues with Bardi choices File

```
verb1_name=verb1
  verb1_feat2_name=case
  verb1_feat2_value=erg
  verb1_feat2_head=subj
verb1_valence=trans
  verb1_stem1_orth=n
  verb1_stem1_pred=_cont_v_rel
verb13_name=verb13
  verb13_feat2_name=case
  verb13_feat2_value=erg
  verb13_feat2_head=subj
verb13_valence=trans
  verb13_stem1_orth=n
  verb13_stem1_pred=_rem.pst_v_rel
```

```
verb-pc1_name=verb-pc1
verb-pc1_order=prefix
verb-pc1_inputs=verb1, verb13
  verb-pc1_lrt1_name=verb-pc1_lrt1
    verb-pc1_lrt1_lri1_inflecting=yes
    verb-pc1_lrt1_lri1_orth=jalala-
    verb-pc1_lrt1_lri2_inflecting=yes
    verb-pc1_lrt1_lri2_orth=rli-
    verb-pc1_lrt1_lri3_inflecting=yes
    verb-pc1_lrt1_lri3_orth=loorroo-
    verb-pc1_lrt1_lri4_inflecting=yes
    verb-pc1_lrt1_lri4_orth=laba-
    verb-pc1_lrt1_lri5_inflecting=yes
    verb-pc1_lrt1_lri5_orth=jayboo-
  ...
```

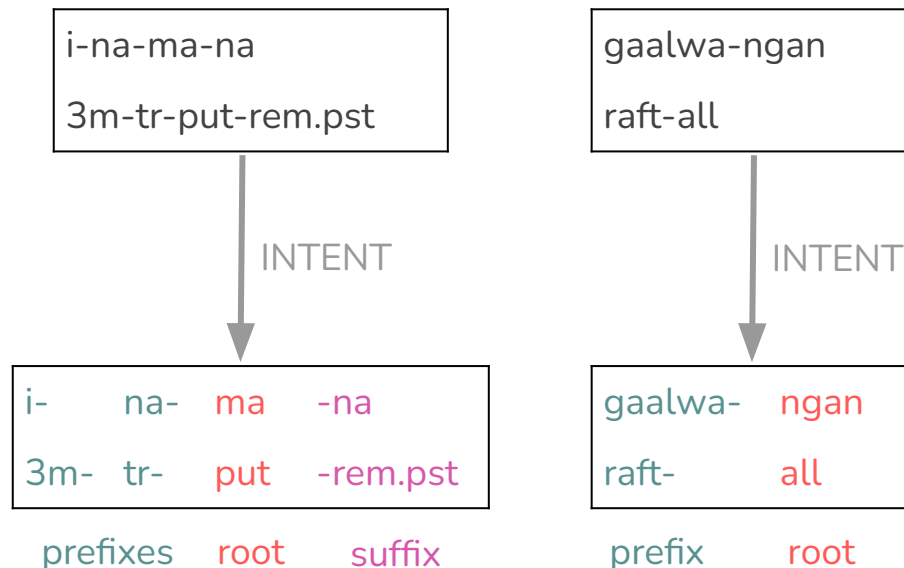
(1) Goolamana barda jarrgany i-na-ma-na gaalwa-ngan.  
G. away cut.across 3M-TR-put-REM.PST raft-ALL  
Goolamana crossed over to his raft. (Bower, 2012, p. 704)





# Stage 1 - Using spacy for verbs

1. Split morpheme into parts (i.e. rem.pst -> [rem, pst] or put -> [put])
2. Use `spacy` to find POS for those parts
3. If one of them is tagged as a verb, mark that morpheme as the root





## Stage 2 - Using Known Grams

1. Keep list of known grams (i.e. 1pl, fut, poss, etc.)
  - Composed of automatically collected grams from ODIN and other grams collected manually
  - Additionally keep track of grams that conflict with English words (i.e. all, pass, sing, etc.)
2. For each word, find the morphemes that are not grams
  - Keep track of any morphemes that are conflict grams
  - Also find verb root with `spacy` as back-up
3. Determine the root



## Stage 2 - Using Known Grams

If at least 1 morpheme in the word is not a gram (is a possible root):

- If more than 1 morpheme is not a gram and a verb root was identified with spacy
  - Find the root using spacy based on the list of non-grams
- If exactly 1 morpheme is not a gram
  - Root is that morpheme

If all morphemes are grams (no possible roots):

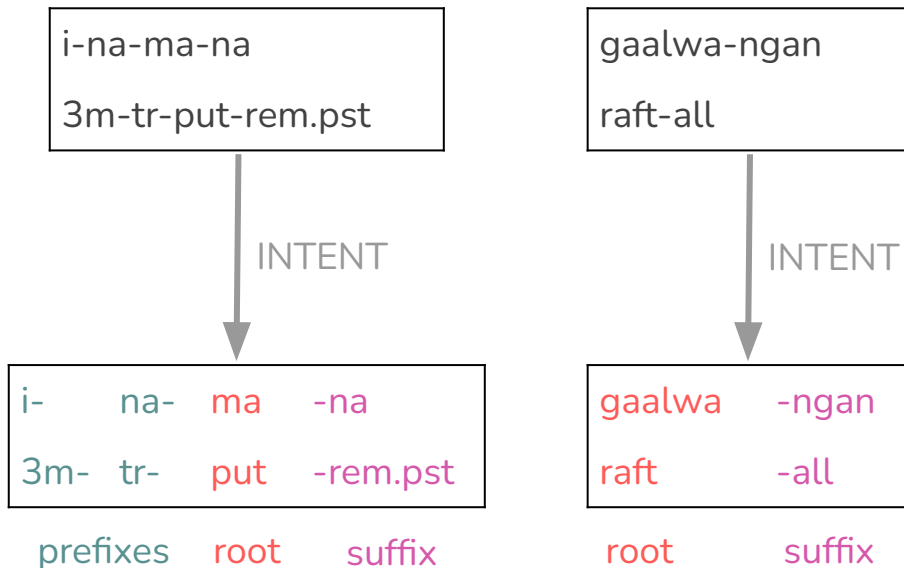
- If exactly 1 conflict gram was identified
  - Root is that gram
- If a verb root was identified with spacy
  - Root is that verb root



# Results

For Bardi:

- (almost) no affixes in lexicon  
`verb34_name=verb34`  
`verb34_valence=trans`  
`verb34_stem1_orth=irr~`  
`verb34_stem1_pred=_3a_v_rel`
- (almost) no verb roots in position classes  
`verb-pc33_name=verb-pc33`  
`verb-pc33_order=suffix`  
`verb-pc33_inputs=verb11`  
`verb-pc33_lrt1_name=verb-pc33_lrt1`  
`verb-pc33_lrt1_lri1_inflecting=yes`  
`verb-pc33_lrt1_lri1_orth=-roowil`





# References

- Bender, Emily M, Dan Flickinger, and Stephan Oepen. 2002. The grammar matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics, pages 8–14, Taipei.
- Bender, Emily M, Scott Drellishak, Antske Fokkens, Laurie Poulson, and Safiyyah Saleem. 2010. Grammar customization. *Research on Language & Computation*, 8(1):23–72. 10.1007/s11168-010-9070-1.
- Bowern, Claire. 2012. A grammar of Bardi, volume 57. Walter de Gruyter.
- Georgi, Ryan. 2016. From Aari to Zulu: Massively Multilingual Creation of Language Tools Using Interlinear Glossed Text. Ph.D. thesis, University of Washington.