Revisiting clitic phenomena in Mixtec

54th Annual Meeting of the Societas Linguistica Europaea
30.08.-03.09.2021

Sandra Auderset [1,2], Adam J.R. Tallman [1,3], Carmen Hernández Martínez [2]

[1] Max Planck Institute for Evolutionary Anthropology
[2] University of California Santa Barbara
[3] Friedrich-Schiller-Universität Jena

31 August 2021

(1) San Martín Duraznos Mixtec

    a. *chíntxeé **ta'an tuun*** ***ntxìvà'-na***
       IMPF.help RECP always INTENS-3PL.HUM
       'They always help each other a lot.'

    b. ***ta sáàn** ìsùvá'-**ì*** ***ìsísínì-va-ì***
       and then PFV-prepare-1SG PFV-have.breakfast-ADD-1SG
       'And then I prepared breakfast.'

Pike's[1] analysis:

1. morphology-syntax distinction is not informative
2. constituency exists but there is no discrete shift in structure from morphology to syntax

Macaulay's[2] analysis:

1. Pike's *deep isolating hypothesis* (that all bound forms are underlyingly free forms) is wrong
2. there is a morphology-syntax distinction
3. clitics are a motivated class

---

[1]Pike, K.L. 1944. Analysis of a Mixteco text. IJAL 10(4); Pike, K.L. 1945. A problem in morphology-syntax division. Acta Linguistica 5(1)

[2]Macaulay, M. 1987. Cliticization and the morphosyntax of Mixtec. IJAL 53(2)

Criteria:

|	No. of syll.	Selection	Pos. wrt. stem
Words	2	NA	NA
FspC	1	various cat.	NA
PhA	1	various cat.	further
Aff	1	only one cat.	closer

Elements discussed:

- *sa-/s-* CAUS (Aff)
- *be-* BUILDING (Aff)
- *ti-* CLF.ANIM (Aff)
- dependent pronouns (PhA)
- independent pronouns (Words)
- *ha-* complementizer (PhA)
- *ha-* deadj. NMLZ (Aff)

1. Are clitics synchronously motivated?
2. What does it mean for clitics to be 'theory-neutral'?
3. Is Macaulay's methodology adequate for substantiating her claims? If not, what methodology is?
4. Is there a morphology-syntax distinction?

Pike's proposal
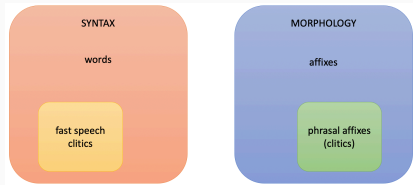
Macaulay's proposal

San Martín Duraznos Mixtec

- 106 closed class elements including things traditionally classified as affixes, clitics, and words (pronouns, conjunctions, aspect/mood markers, intensifiers, classifiers, temporal expressions etc.)
- forms collected from a 7h corpus of naturalistic speech complemented with elicitation and native speaker judgements by co-author Carmen Hernández
- each element is coded for an array of diagnostics seen as crucial for distinguishing affixes/clitics/words (mostly based on Spencer & Luís[3]) following the methdology of multivariate typology[4]

---

[3] Spencer, A. & A.R. Luís. 2012. Clitics: an introduction. Cambridge University Press
[4] Bickel et.al. 2011. Multivariate typology and field linguistics: a case study on detransitivization in Kiranti (Sino-Tibetan). Proceedings of the Conference on Language Documentation and Linguistic Theory 3

Variable	Description	Values
Selection	How many categories (of V, N, A) does it combine with?	1,2,3
Permutability	Is it permutable with the head?	no/yes/diff
Permutabilty Dist	Is it permutable with distributionally identical elements?	no/yes/diff
Interruptability	Is it interruptable from the head?	no/bound/free
Boundedness	Can it stand alone as a complete utterance?	yes/related/no
Morae	How many morae does it have?	1,2,3
Allomorphy	Is there allomorphy?	none/low/high
Wide Scope	Does it have wide scope under coordination?	yes/no
Internal Complexity	Is it composed of identifiable parts (morphs)?	yes/no
Fossilization	Does it combine with non-recurring bases?	yes/no

Hierarchical clustering

- for assessing whether a two-way partition is motivated
- compared to a baseline: sampled from data and sampled from a uniform distribution
- ratio of the first partition to total cluster height as an indication of how well supported it is by the data

Random Forest

- for assessing which and how many variables are important in the classifications
- overall error indicates how well a classification does
- comparison with the baseline (proportion of the largest class) shows how much the classification adds
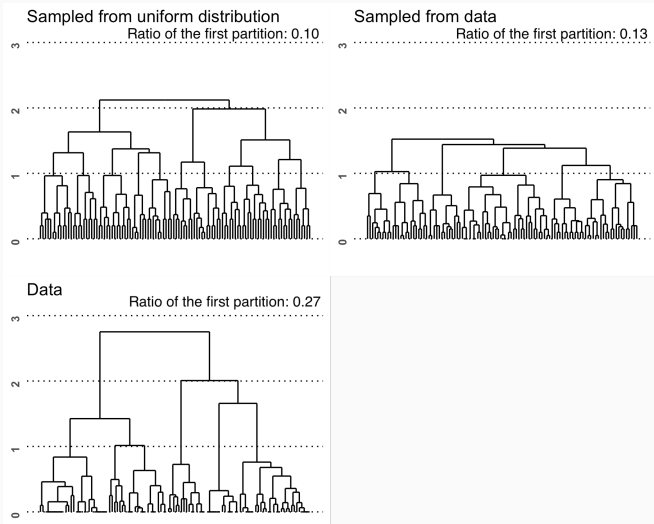
Hierarchical clustering of the data overlayed with Macaulay's classification
(blue=Aff, green=PhAff, orange=W)

Hierarchical clustering of the data compared with two simulated data sets, one sampled from the data and one sampled from a uniform distribution

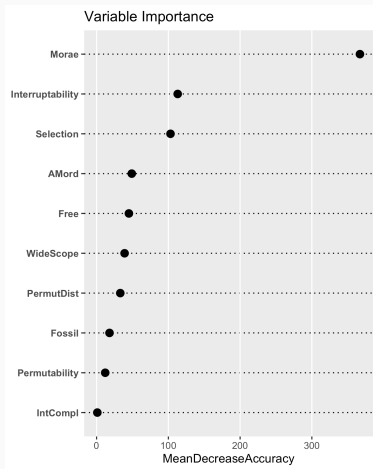Random Forest with Macaulay's classification as the dependent variable and all others as independent variables

baseline	0.68
accuracy	0.85
difference	+0.17

```
        OOB estimate of  error rate: 4.72%
Confusion matrix:
        Affix PhAffix Word class.error
Affix      7       3    0  0.30000000
PhAffix    1      23    0  0.04166667
Word       0       1   71  0.01388889
```
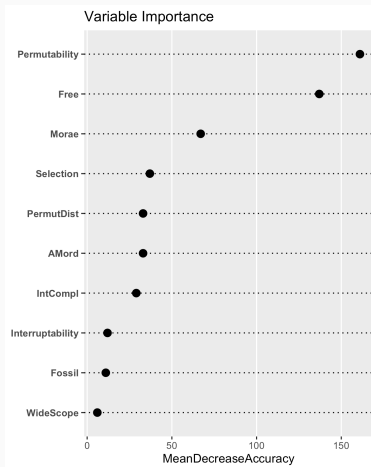

Variable Importance

Random Forest with the first partition of the cluster as the dependent variable and all others as independent variables

baseline	0.53
accuracy	0.98
difference	+0.45

```
        OOB estimate of  error rate: 1.89%
Confusion matrix:
   1  2 class.error
1 54  2  0.03571429
2  0 50  0.00000000
```

Variable Importance

Is a clitic category motivated?

Random Forest with the second partition of the cluster as the dependent variable and all others as independent variables

baseline	0.72
accuracy	0.98
difference	+0.26

```
         OOB estimate of  error rate: 2.83%
Confusion matrix:
   1  2  3 class.error
1 39  0  2  0.04878049
2  1 14  0  0.06666667
3  0  0 50  0.00000000
```


Variable Importance of RF with second cluster
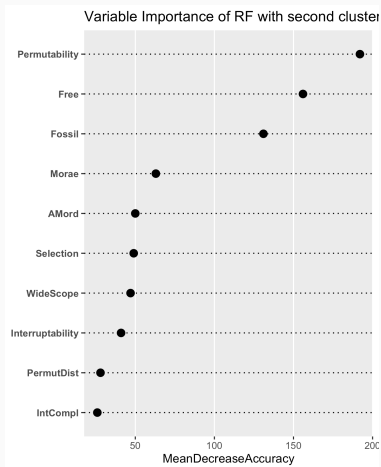
Pike's analysis:

- an *a priori* postulated morphology-syntax distinction is not meaningful BUT some partition can be motivated
- investigate actual correlates between properties

Macaulay's analysis

- clitics are not a motivated category
- morphology-syntax distinction motivated on an abstract level BUT not based on her criteria
- methodology that tautologically leads to the number of classes she postulates
- clitics cannot exist apart from a theory
- no textual evidence for Macaulay's interpretation of Pike

1. Are clitics synchronously motivated? - NO
2. Is there a morphology-syntax distinction? - MAYBE
3. Is there language internal evidence for Macaulay's classification (i.e. is it 'theory-neutral')? - NO

Pike's proposal
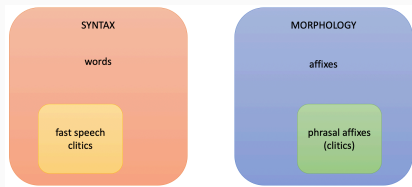
Macaulay's proposal

Acknowledgments

Many thanks to:

Catalina Martínez Ramirez

Reina Martínez Rendón

Pedro Pérez Mendoza
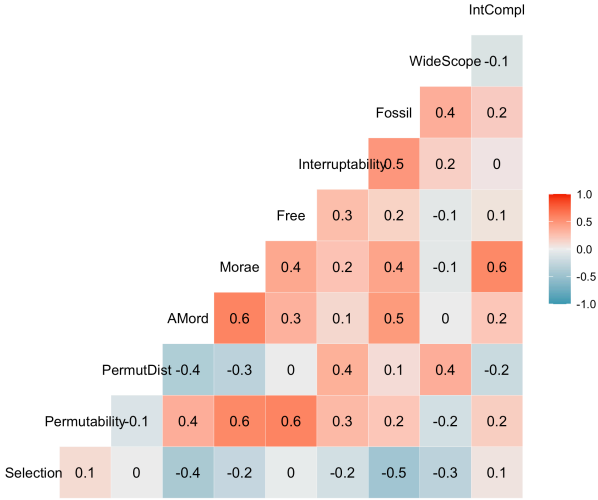
Eric Campbell

Albert Ventayol-Boada

NSF (National Science Foundation)

MICOP (Mixteco/Indígena Community Organizing Project, Oxnard California)

ELDP (Endangered Languages Documentation Programme)

for making this work possible!

Correlation matrix of the variables

Variables and values:

- aggregated: How many classes (out of V, N, A) does the element combine with? - 1, 2, 3
    - Does the element combine with verbs? – yes/no
    - Does the element combine with nouns? – yes/no
    - Does the element combine with adjectives? – yes/no

Assumptions:

- affixes are ciscategorial (combine with one category only)
- clitics and words are transcategorial (combine with multiple categories)

Examples of transcategorial *sá* CAUS:

Form	Gloss	Base	Word class of base
sá-tátán	cure sb.	*tàtàn* 'medicine'	noun
sá-koyó	dekernel	*koyo* 'empty'	verb
sá-ntoo	clean sth.	*ntoo* 'clean'	adjective

Variables and values:

- aggregated: Is the element permutable with the head? - yes/yes with a scope difference/no
  - Can the element variably order with V? – yes/yes with a scope difference/no/NA
  - Can the element variably order with N? – yes/yes with a scope difference/no/NA
  - Can the element variably order with A? – yes/yes with a scope difference/no/NA

Assumptions:

- affixes have a fixed order
- clitics and words have a free(er) order

Examples:

lo'o 'small, little' can occur before and after a verb without scope difference
  ìxinu lo'o-rà 'he ran a little bit'
  lo'o ìxinu-rà 'he ran a little bit'

ta'an RECP has a fixed order
  chíntxeé ta'an-na 'they are helping each other'
  * *ta'an chíntxeé na* intended: 'they are helping each other'

Variables and values:

- Is the element permutable with elements of the same distributional class? - yes/yes with a scope difference/no

Assumptions:

- affixes have a fixed order
- clitics and words have a free(er) order

Example of permutable adverbials *xàà* 'already' and *sa'a* 'like this':

(2) a. *taa ikán   xàà   sa'a-va   ntáa míí iti-nà   ikán*
         and DEM.PROX already like.that-ADD be TOP cornfield-3PL.HUM DEM.PROX
         'and here, their cornfield is already like this here'

     b. *taa ikán   sa'a xàà-va   ntáa mii iti-nà   ikán*
         and DEM.PROX already like.that-ADD be TOP cornfield-3PL.HUM DEM.PROX
         'and here, their cornfield is already like this here'

Allomorphy

Variables and values:

- aggegreated: Does the element exhibit allomorphy and if so, to what degree? – none, low (2-4), high (>4)
  - How many segmental allomorphs does the element have? – 1-n
  - Does this allomorphy pattern appear with other elements of the same phonological structure? – unique, limited, various, general
  - How many tonal allomorphs does the element have? – 1-n
  - What is the distribution of the allomorphy pattern? – unique, limited, various, general
  - Allomorphy score: (segmental x tonal) + distribution adjustments [general=0, various=1, limited=2, unique=3]

Assumptions:

- affixes exhibit high degrees of allomorphy, especially of irregular allomorphy
- words exhibit no allomorphy or only regular allomorphy
- clitics fall somewhere between

Examples:

- 2sg.nhon dependent pronoun has high allomorphy (score of 6) because it has three allomorphs: *ùn*, *un*, and *ún* and this allomorphy is limited to this specific pronoun
- conjunction 'but' has low allomorphy (score of 3) because it has two allomorphs: *soo* and *suu* and this alternation is observed is many other items

Variables and values:

- Does the element combine with non-recurring bases? – yes/no

Assumptions:

- affixes can combine with non-recurring/fossilized bases
- words and clitics do not combine with non-recurring bases

Example: *na* PL.HUM combining with a recurring and a non-recurring base
ná-ñuu 'community' from: *ñuu* 'village'
nà-yivi 'people' (*yivi* does not occur by itself or in other combinations)

Variables and values:

- How many morae does the element have? – 1, 2, 3

Assumptions:

- words are always bimoraic (or more)
- affixes and clitics are monomoraic

Examples: PFV marker *ì* has 1 mora, question word *ntxáa* 'where?' has 2 morae

Free occurrence

Variables and values:

- Can the element be used in isolation as a complete utterance as is? – yes/in a related form/no
- Note: this is not the same as bimoraicity - lots of bimoraic forms cannot be used as complete utterances

Assumptions:

- affixes can never be used as free forms
- words can always be used as free forms
- clitics can be used as free forms as is or in a related form, depending on their class

Examples:

- the masculine classifier *txà* cannot be used as a free as is, but has a related free from *txàa* 'man' that is a free form
  - *txà-kuì'ná* 'thief (m)' [CLF.M-steal]
  - **txà* intended: 'a man/male person'
  - *txàa* '(a) man (e.g. as a response to "who's over there?")
- the classifier *txí* for spherical things cannot be used as a free form and does not have a related free form
  - *txí-vìshì* 'candy' [CLF.SPH-sweet]
  - **txí* intended 'a spherical thing'

Variables and values:

- aggregated: Can the element be interrupted from the head/host? – no/by a bound form/by a free form/by a bound or a free form
- Can the element be interrupted from the head/host by a free form? – yes/no
- Can the element be interrupted from the head/host by a bound form? – yes/no

Assumptions:

- affixes cannot be interrupted from their head or only by a bound form
- clitics and words can be interrupted by free forms

Example: the irrealis negation *u/o~i* can be interrupted from the verb by a bound form, but not by a free form

(3) a. *u-ntá-ntikí-ún*
        NEG.IRR-ITER-look.for-2SG.NHON
        'You won't be searching for it again.'

   b. * *u   ntáá ntá-ntikí-ún*
        NEG.IRR truly ITER-look.for-2SG.NHON
        intended: 'You surely won't be searching for it again.'

Variables and values:

- Can the element have scope over coordinated hosts? – yes/no

Assumptions:

- affixes cannot have scope over coordinated heads
- clitics can have scope over coordinated elements

Example: the perfective marker *ì* cannot have scope over two verbs

(4) a. *ta   sáàn ì-sùvá'-ì   ì-sísínì-va-ì*
         and then PFV-prepare-1SG PFV-have.breakfast-ADD-1SG
         'and then I prepared breakfast'

     b. **ta sáàn ì-sùvá'-ì sísínì-va-ì* intended: 'and then I prepared breakfast'

Variables and values:

- Does the element have internal complexity (morphs)? – yes/no

Assumptions:

- affixes and clitics do not have internal complexity
- words can have internal complexity

Examples:

yì'î	1SG	independent pronoun	no internal complexity
ntxìvà'a	INTENSIFIER		composed of the morph *ntxì* and the morpheme *và'a* 'good, well'