



University of
Zurich ^{UZH}

Romanisches Seminar

A baseline for object clitic climbing in Italian—preliminary results

Francesco Gardani & Chiara Zanini

RoLinC (Romance Linguistics Circle), 5 March 2024



Roadmap

- The Phenomenon
- State-of-the-Art and Goal
- Study's Design
- Data and Results
- Conclusion



The Phenomenon

(1) Italian

‘Marco wants to buy it.’



The Phenomenon

(1) Italian

a.	<i>Marco</i>	<i>vuole</i>	<i>comprar=lo</i>
	Marco	wants	buy.INF=DO

‘Marco wants to buy it.’



The Phenomenon

- (1) Italian
- a. *Marco vuole comprare=lo*
 Marco wants buy.INF=DO
- b. *Marco lo=vuole comprare*
 Marco DO=wants buy.INF
 ‘Marco wants to buy it.’



The Phenomenon

- (1) Italian
- | | | | |
|----|---------------------|------------------------|------------------------|
| a. | <i>Marco</i> | <i>vuole</i> | <i>comprar=lo</i> |
| | Marco | wants | buy.INF=DO |
| b. | <i>Marco</i> | <i>lo=vuole</i> | <i>comprare</i> |
| | Marco | DO=wants | buy.INF |
- ‘Marco wants to buy it.’

Object Clitic Climbing (OCC) in (1b): a clitic pronoun realizing a direct object (DO) selected by a complement verb moves out of its local domain and climbs to the matrix clause, in the presence of restructuring verbs.



The Phenomenon

- (1) Italian
- | | | | |
|----|--------------|-----------------|-------------------|
| a. | <i>Marco</i> | <i>vuole</i> | <i>comprar=lo</i> |
| | Marco | wants | buy.INF=DO |
| b. | <i>Marco</i> | <i>lo=vuole</i> | <i>comprare</i> |
| | Marco | DO=wants | buy.INF |
- ‘Marco wants to buy it.’

comparison between (1a) and (1b):

- enclisis (1a) vs proclisis (1b)
- same interpretation, both semantically and pragmatically



State-of-the-Art and Goal

“The question of which combination permits or requires clitic climbing is also subject to variation in Romance, with particular language-specific and construction-specific options” (Dragomirescu et al. 2022: 574)



State-of-the-Art and Goal

“The question of which combination permits or requires clitic climbing is also subject to variation in Romance, with particular language-specific and construction-specific options” (Dragomirescu et al. 2022: 574):

- OCC in Italo-Romance varieties (IRVs) (Benincà 1986; Jones 1993; Ledgeway 1998, 2000; Manzini & Savoia 2005; Monachesi 1995)
- diachronic evolution of OCC (cf. Pescarini 2021: 235–266)
- syntactic contexts licensing OCC (cf. Cardinaletti & Giusti 2022; Cardinaletti & Shlonsky 2004: 574)



State-of-the-Art and Goal

What do we know about OCC in Italian?



State-of-the-Art and Goal

What do we know about OCC in Italian?

- “in Italian, clitic climbing is optional” (Monachesi 1995: 6, fn4)
- “*Carlo lo andrà a comprare* è più usata nell’italiano toscano e centro-meridionale; *Carlo andrà a comprarlo* è più usata in quello settentrionale” (Calabrese 2001: 586)



State-of-the-Art and Goal

What do we know about OCC in Italian?

- “in Italian, clitic climbing is optional” (Monachesi 1995: 6, fn4)
- “*Carlo lo andrà a comprare* è più usata nell’italiano toscano e centro-meridionale; *Carlo andrà a comprarlo* è più usata in quello settentrionale” (Calabrese 2001: 586)

Goal: to fill this massive descriptive lacuna, detecting a robust baseline for OCC in (standard) Italian



State-of-the-Art and Goal

What do we know about OCC in Italian?

- “in Italian, clitic climbing is optional” (Monachesi 1995: 6, fn4)
- “*Carlo lo andrà a comprare* è più usata nell’italiano toscano e centro-meridionale; *Carlo andrà a comprarlo* è più usata in quello settentrionale” (Calabrese 2001: 586)

Goal: to fill this massive descriptive lacuna, detecting a robust baseline for OCC in (standard) Italian

Note that ongoing work by Cardinaletti, Giusti, and Lebani was presented at 16th and 17th CIDSMS (Cardinaletti 2023; Cardinaletti, Giusti & Lebani 2022). Data and results of these studies are not public yet.



State-of-the-Art and Goal

What do we know about OCC in Italian?

- “in Italian, clitic climbing is optional” (Monachesi 1995: 6, fn4)
- “*Carlo lo andrà a comprare* è più usata nell’italiano toscano e centro-meridionale; *Carlo andrà a comprarlo* è più usata in quello settentrionale” (Calabrese 2001: 586)

Goal: to fill this massive descriptive lacuna, detecting a robust baseline for OCC in (standard) Italian

Note that ongoing work by Cardinaletti, Giusti, and Lebani was presented at 16th and 17th CIDSIM (Cardinaletti 2023; Cardinaletti, Giusti & Lebani 2022). Data and results of these studies are not public yet.

Context of discovery: cf. Gardani's 2019 project at Columbia University (NYC):

<https://italianacademy.columbia.edu/directory/francesco-gardani> >

https://italianacademy.columbia.edu/sites/default/files/content/GARDANI_Francesco_Project.pdf



Study's Design

- developed questionnaire:
 - 11 questions targeting the informant's sociolinguistic profile
 - production task: 144 stimuli (48 experimental stimuli and 96 distractors)



Study's Design

- developed questionnaire:
 - 11 questions targeting the informant's sociolinguistic profile
 - production task: 144 stimuli (48 experimental stimuli and 96 distractors): informants presented with stimulus, e.g., "**La macchina non funziona. [devo; riparar(e)]**"



Study's Design

- developed questionnaire:
 - 11 questions targeting the informant's sociolinguistic profile
 - production task: 144 stimuli (48 experimental stimuli and 96 distractors): informants presented with stimulus, e.g., "La macchina non funziona. [devo; riparar(e)]" and asked to build a sentence along the following steps:
 1. first, read the stimulus (e.g., *La MACCHINA non funziona*);
 2. then, read the words in square brackets (e.g., *devo; riparar(e)*);
 3. now, create a sentence that continues the stimulus by:
 - using all the words contained in square brackets; and
 - not repeating the words in caps (e.g., *MACCHINA*) nor replacing them with similar words (e.g. *auto*).



Study's Design

- developed questionnaire:
 - 11 questions targeting the informant's sociolinguistic profile
 - production task: 144 stimuli (48 experimental stimuli and 96 distractors): informants presented with stimulus, e.g., "La macchina non funziona. [devo; riparar(e)]" and asked to build a sentence along the following steps:
 1. first, read the stimulus (e.g., *La MACCHINA non funziona*);
 2. then, read the words in square brackets (e.g., *devo; riparar(e)*);
 3. now, create a sentence that continues the stimulus by:
 - using all the words contained in square brackets; and
 - not repeating the words in caps (e.g., *MACCHINA*) nor replacing them with similar words (e.g. *auto*).

target: either *devo ripararla* or *la devo riparare*



Study's Design

- developed questionnaire:
 - dependent variables: Enclisis, Proclisis, Other
 - independent variables:
 - sociolinguistic
 - linguistic



Study's Design

- developed questionnaire:
 - dependent variables: Enclisis, Proclisis, Other
 - independent variables:
 - sociolinguistic:
 - Age_group: generation (baby boomers, generation X, generation Y (Millennials), generation Z)
 - PoL: place of living (N(orth), C(enter), U(pper)S(outh), E(xtreme)S(outh))
 - PoL.14: place of living before age 14 (N(orth), C(enter), U(pper)S(outh), E(xtreme)S(outh))
 - Parent1: origin of parent1/mom (N(orth), C(enter), U(pper)S(outh), E(xtreme)S(outh), other)
 - Parent2: origin of parent2/dad (N(orth), C(enter), U(pper)S(outh), E(xtreme)S(outh)), other)
 - Dialect_Use: frequency of use of IRV (everyday, once a week, three times a week, rarely, never)
 - Dialect: geographic area of IRV (N(orth), C(enter), U(pper)S(outh), E(xtreme)S(outh), n(ot)r(elevant))
 - linguistic



Study's Design

- developed questionnaire:
 - dependent variables: Enclisis, Proclisis, Other
 - independent variables:
 - sociolinguistic
 - linguistic:
 - Verb_Type: modal (*dovere*), motion (*andare a*), aspectual (*finire di*), conative (*provare a*)
 - CL_Num: number value of the clitic (SG vs PL)
 - CL_Gen: gender value of the clitic (M vs F)
 - CL_Ref: animacy value of the clitic's referent (AN vs INAN)
 - Subj_Type: type of subject (NP vs null) in the clause



Study's Design

- developed questionnaire
- administered questionnaire online via the *LimeSurvey* app
- collected data from over 300 informants:
 - Italian as L1
 - resident in Italy (homogeneously distributed across the country) and Italophone Switzerland (Ticino and Grisons)
- thus far (for the purpose of the present presentation) analyzed 238 questionnaires
- conducted analysis by means of the R software (R Core Team 2020), conditional inference trees and random forest models



Data and Results

sample (n = 238):

– place of living:

- N: 113 (48%)
- C: 51 (21.5%)
- US: 47 (20%)
- ES: 26 (10.5%)

– place of living before age 14:

- N: 101 (42%)
- C: 43 (18%)
- US: 62 (26%)
- ES: 32 (14%)

– age group:

- baby boomers (1945-1965): 6 (2%)
- gen X (1966-1980): 34 (14%)
- gen Y (1981-1995): 104 (44%)
- gen Z (1996-2010): 94 (40%)



Data and Results

dependent variables:

- enclisis: 9'245 (81%)
- proclisis: 1'544 (14%)
- other: 635 (5%)



Data and Results

dependent variables:

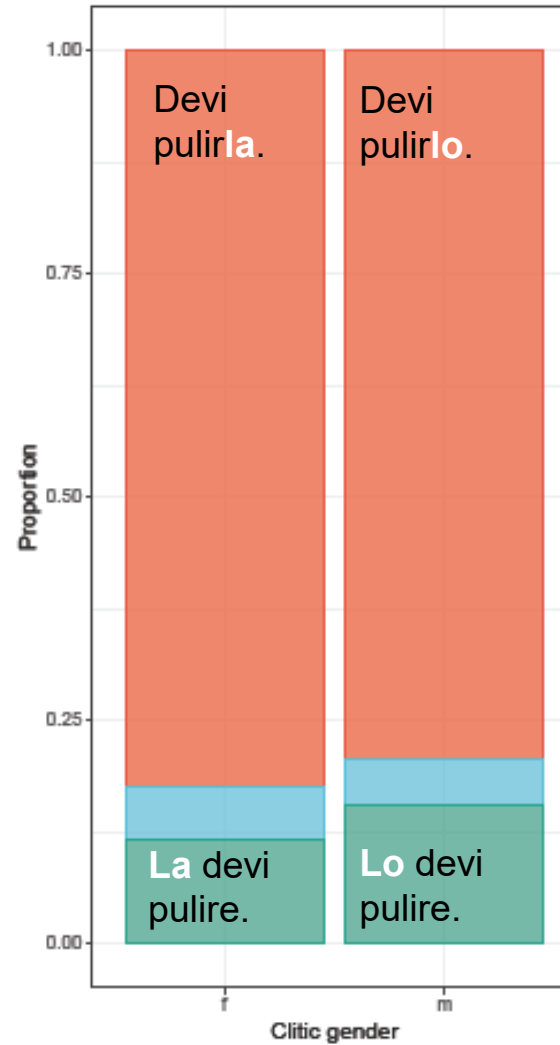
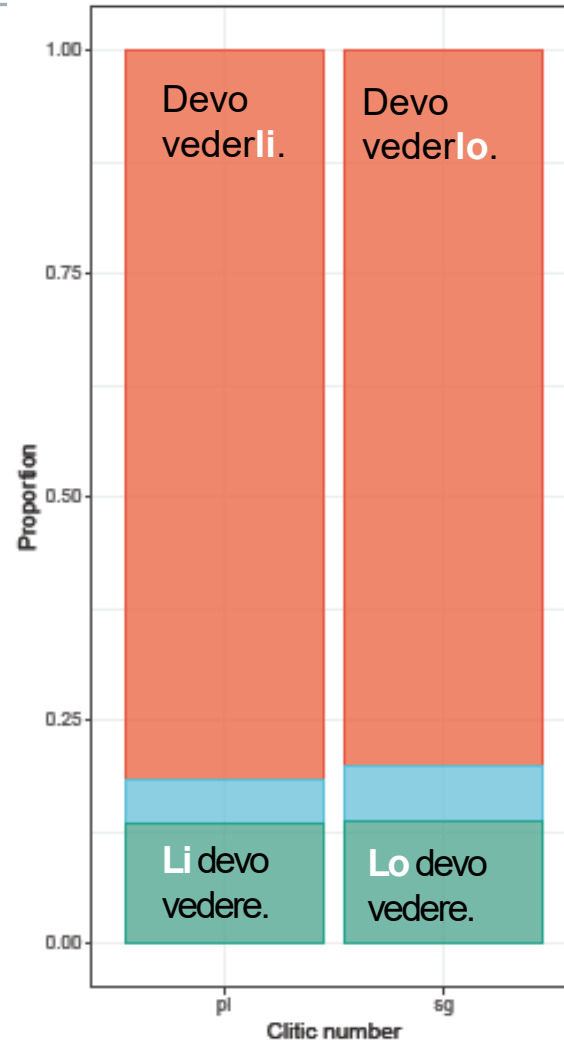
- enclisis: 9'245 (81%)
- proclisis: 1'544 (14%)
- other: 635 (5%)

confirms Monachesi's (1995: 6, fn4) claim: “in Italian, clitic climbing is optional”



Data and Results

number value
(SG vs PL)

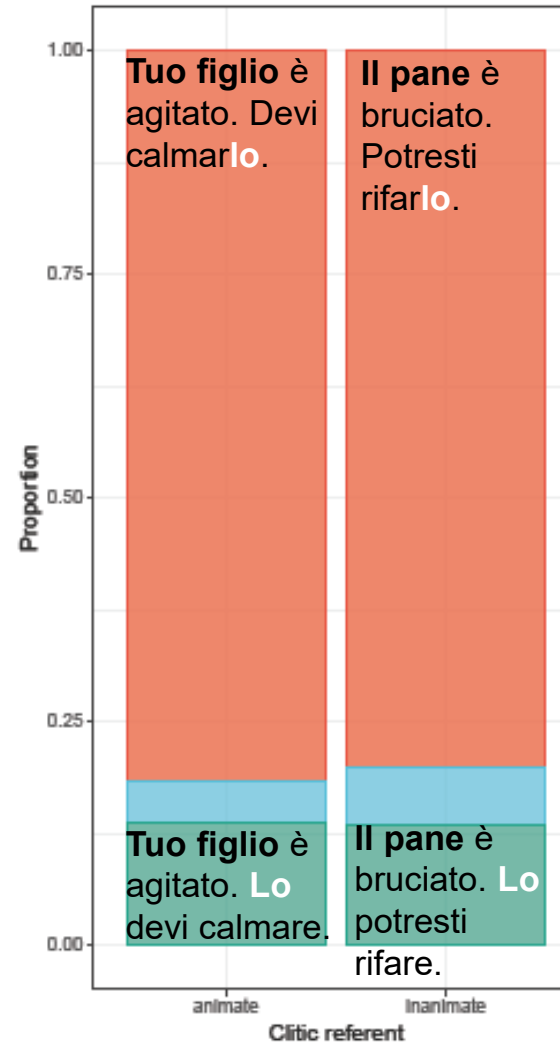
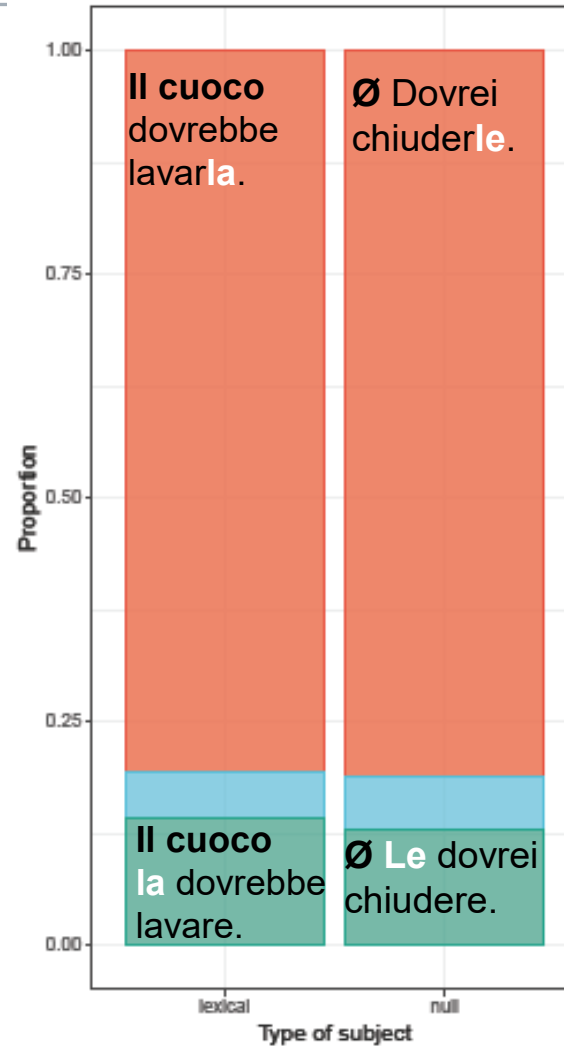


gender value
(M vs F)



Data and Results

type of subject
(NP vs null)

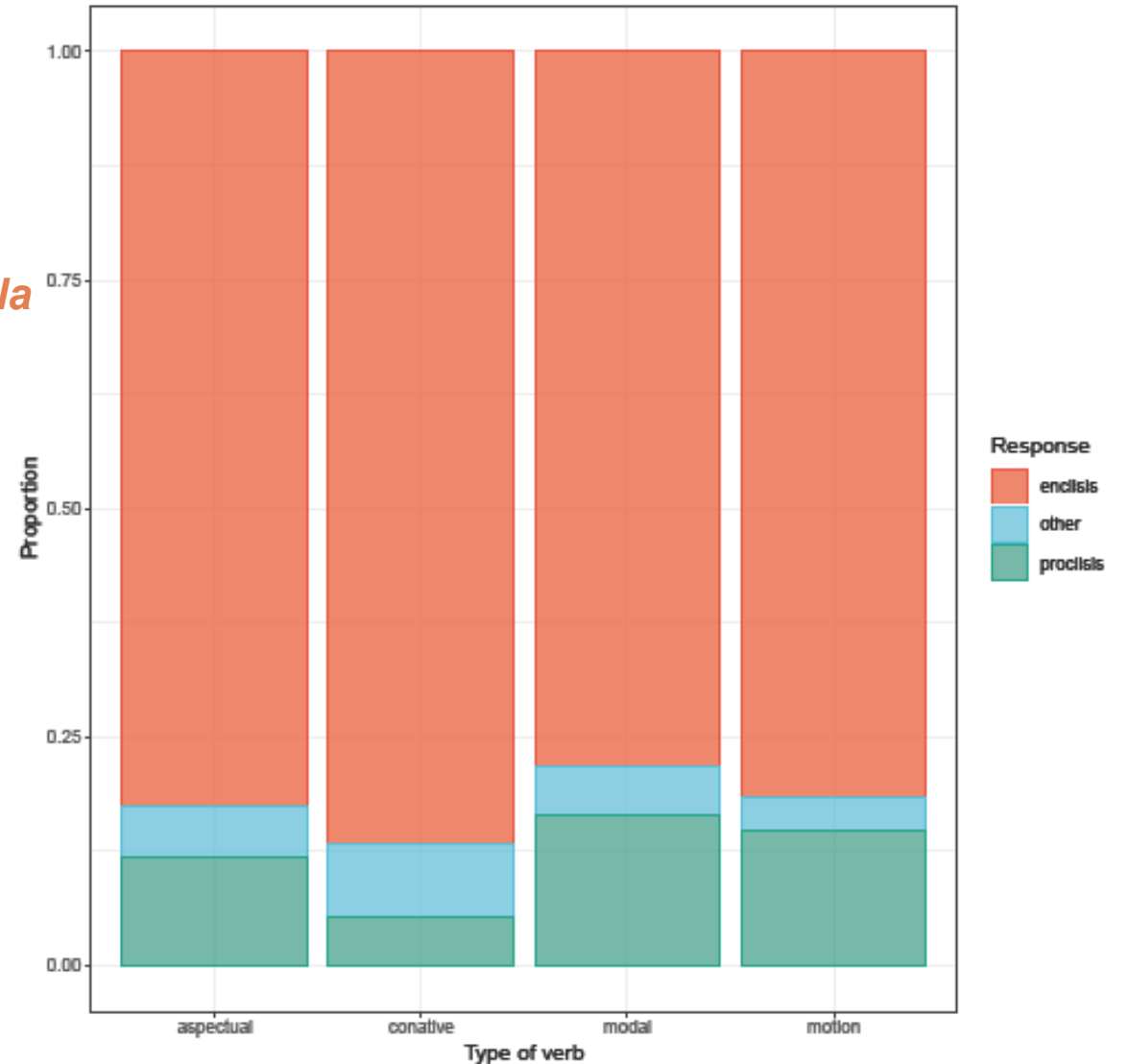


animacy of the clitic's
referent (AN vs INAN)

Data and Results

type of verb:

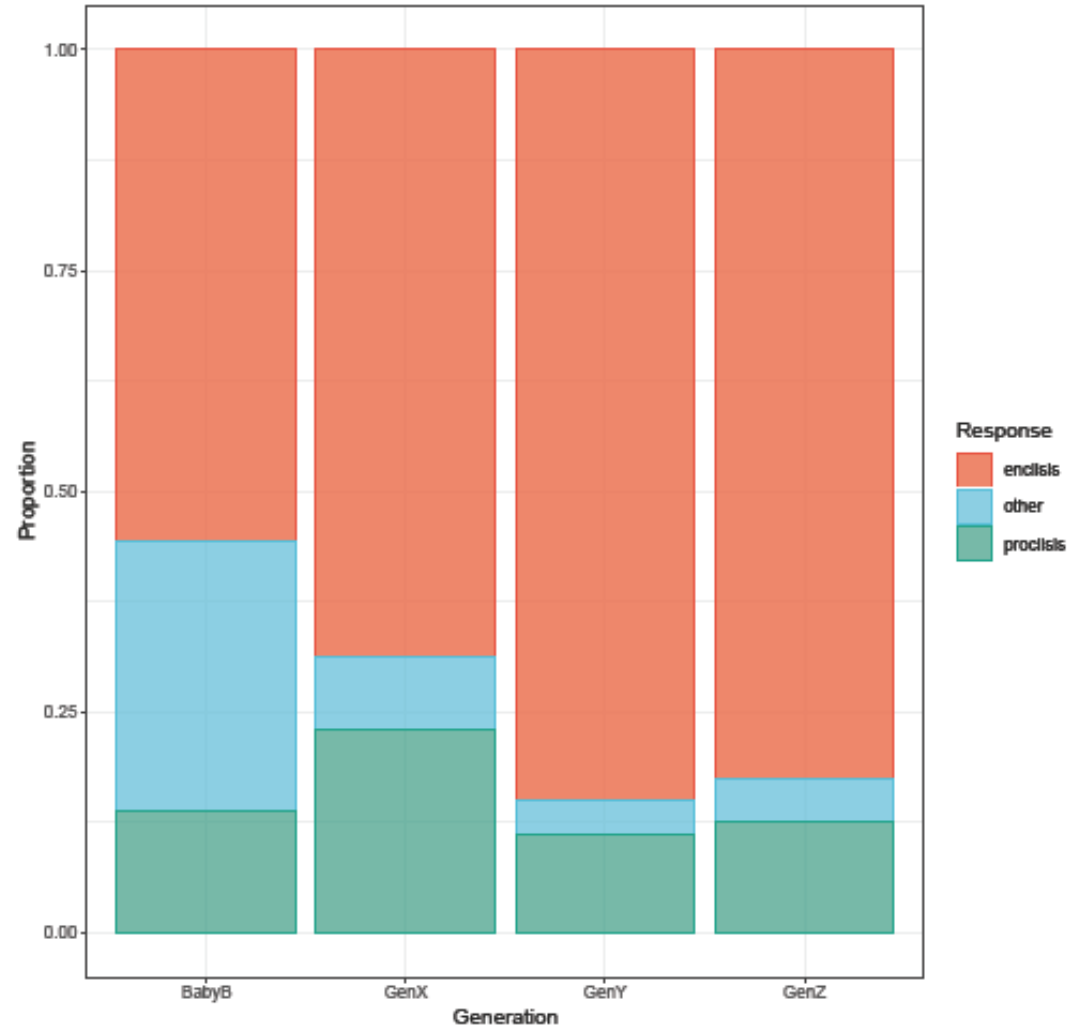
- aspectual: *la sta per presentare / sta per presentarla*
- conative: *la proverà a salvare / proverà a salvarla*
- modal: *la devo riparare / devo ripararla*
- motion: *la torna a trovare / torna a trovarla*





Data and Results

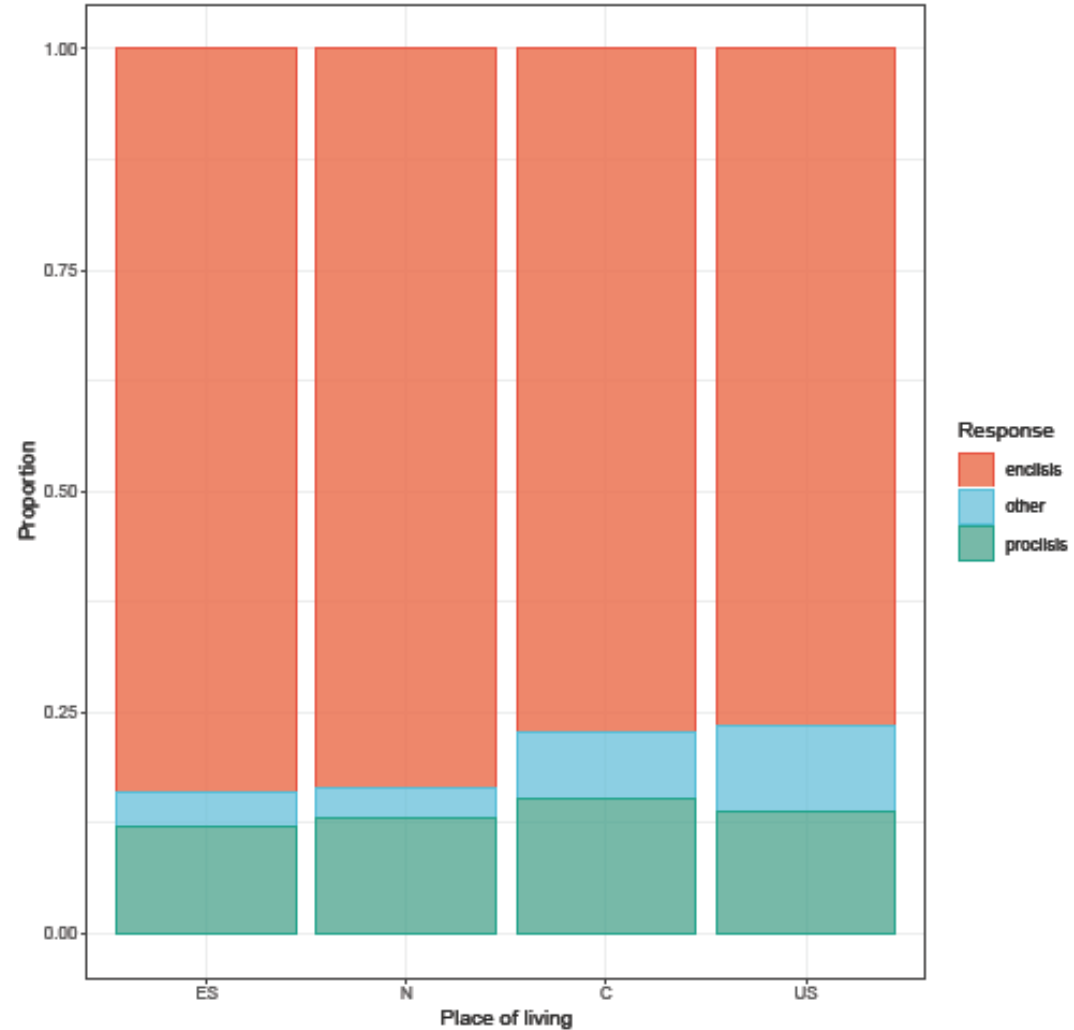
age group:





Data and Results

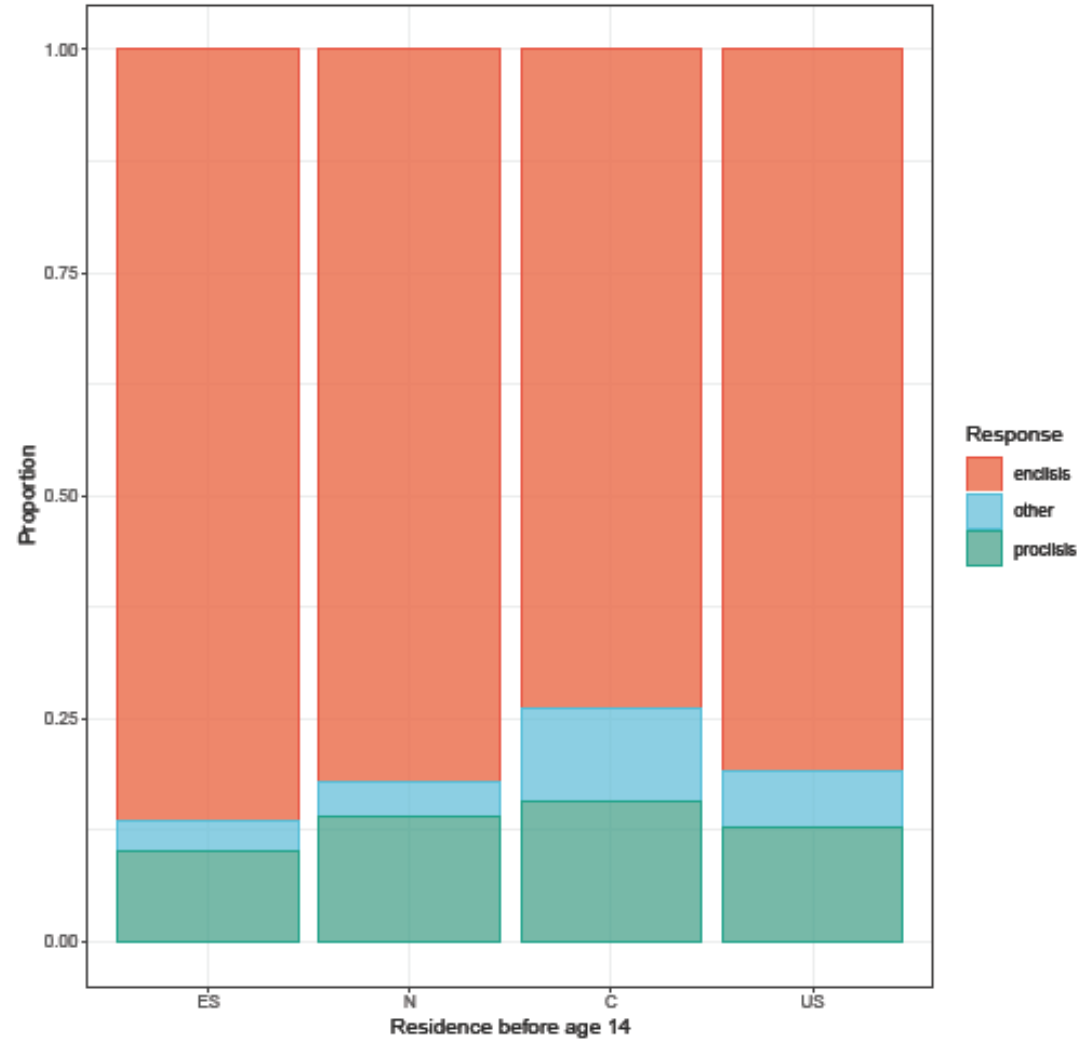
place of living:





Data and Results

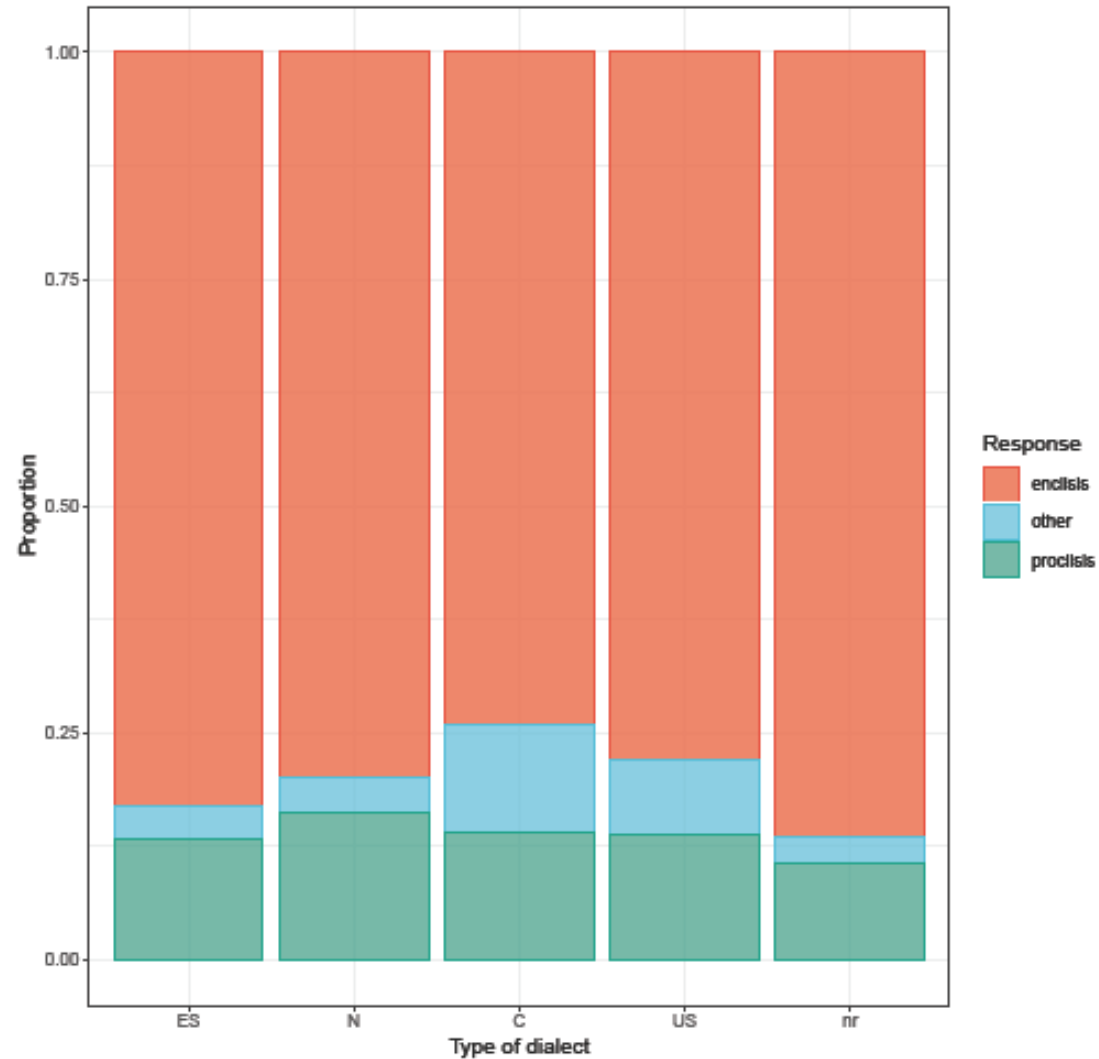
place of living before age 14:





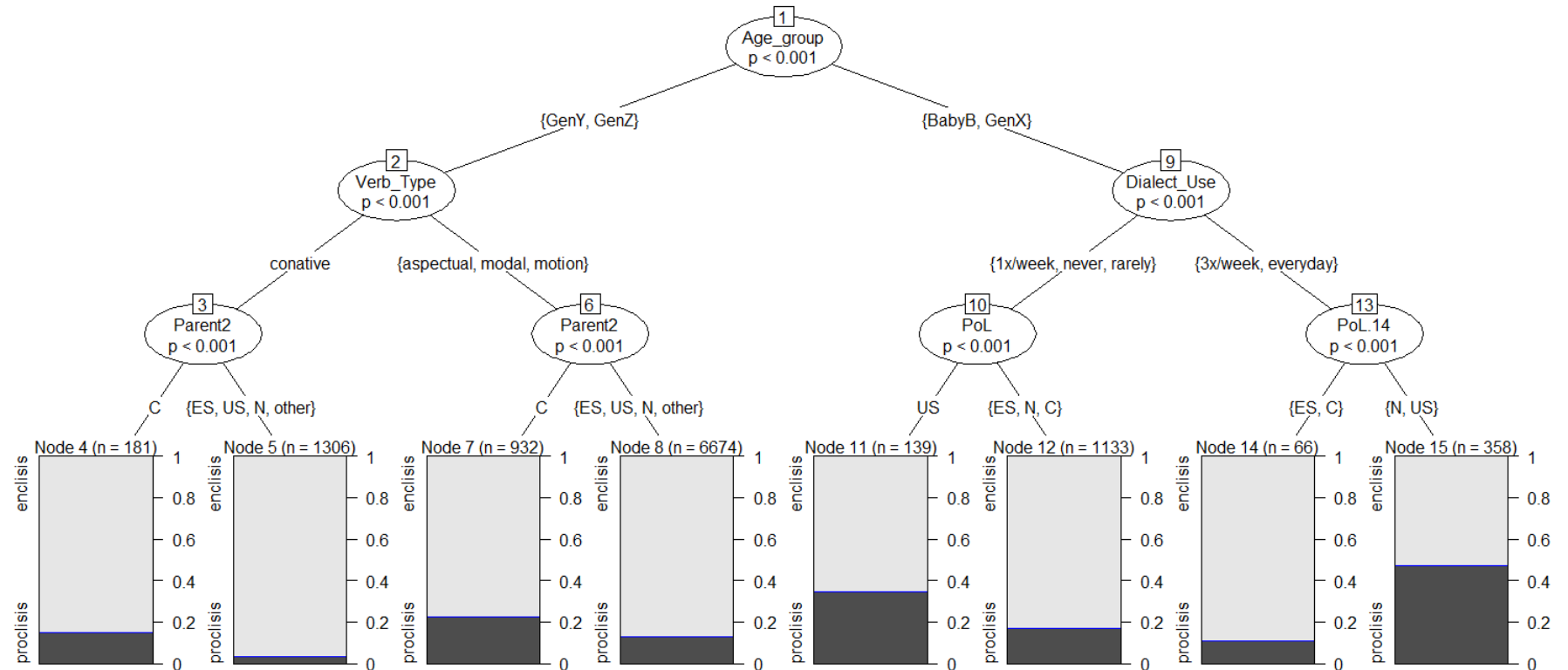
Data and Results

dialect (geographic area of IRV):



Data and Results

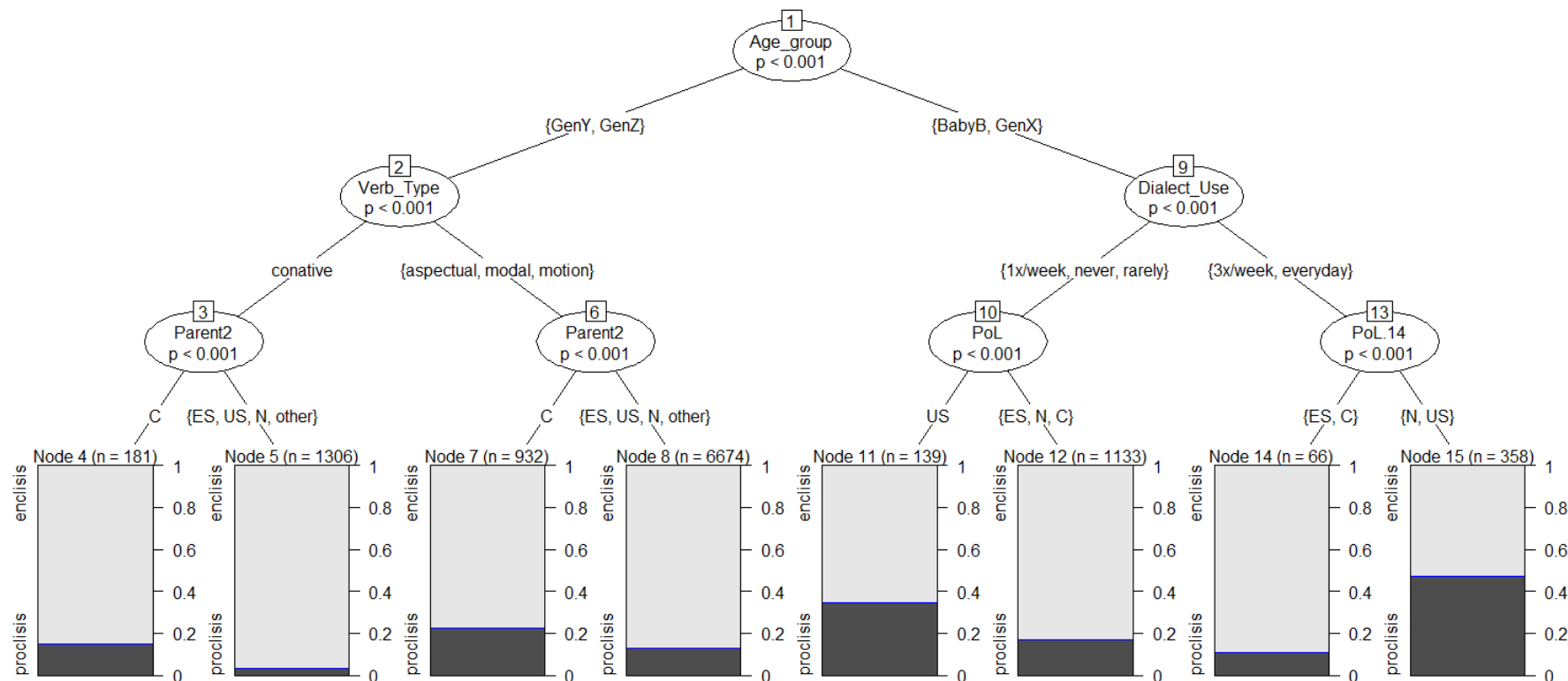
Conditional Inference Tree:



Data and Results

Conditional Inference Tree:

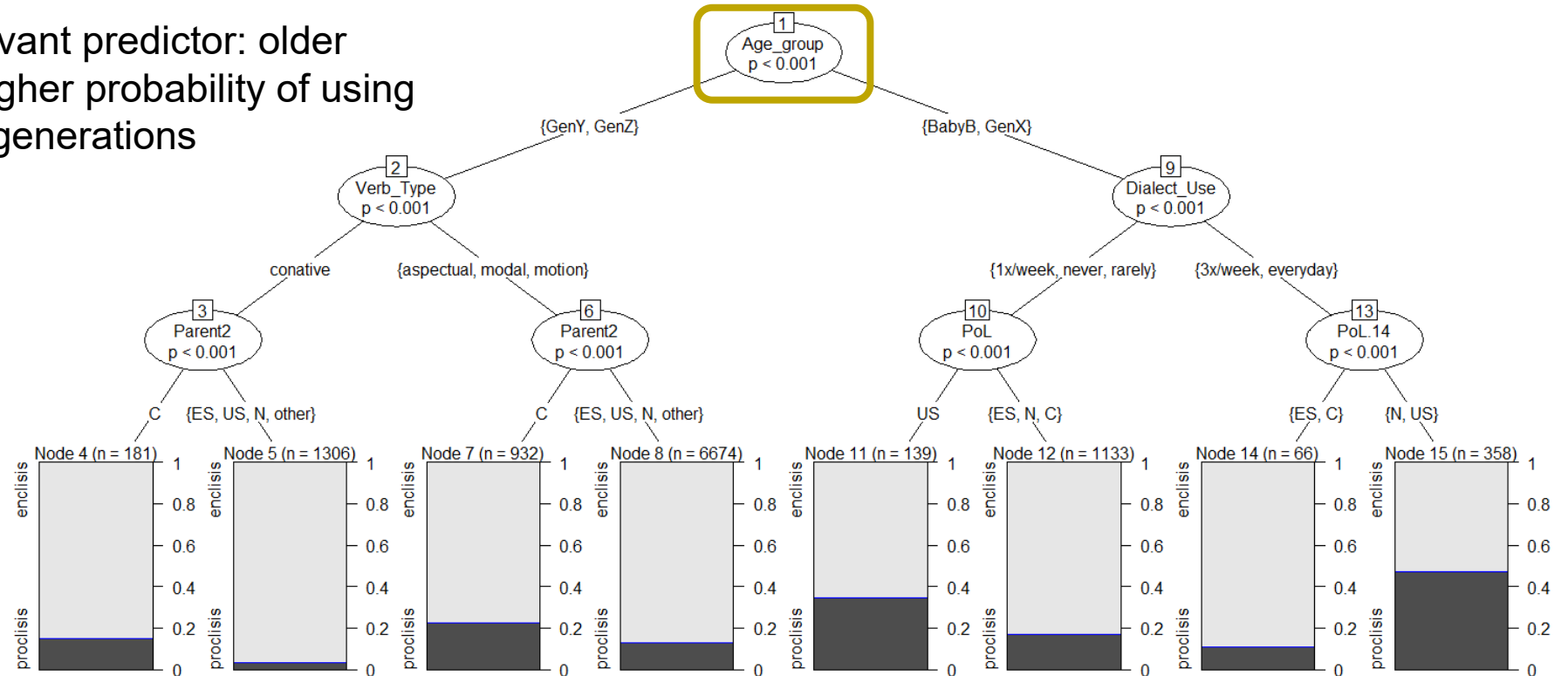
- $C = 0.79$
- reproduced only first three splits



Data and Results

Conditional Inference Tree:

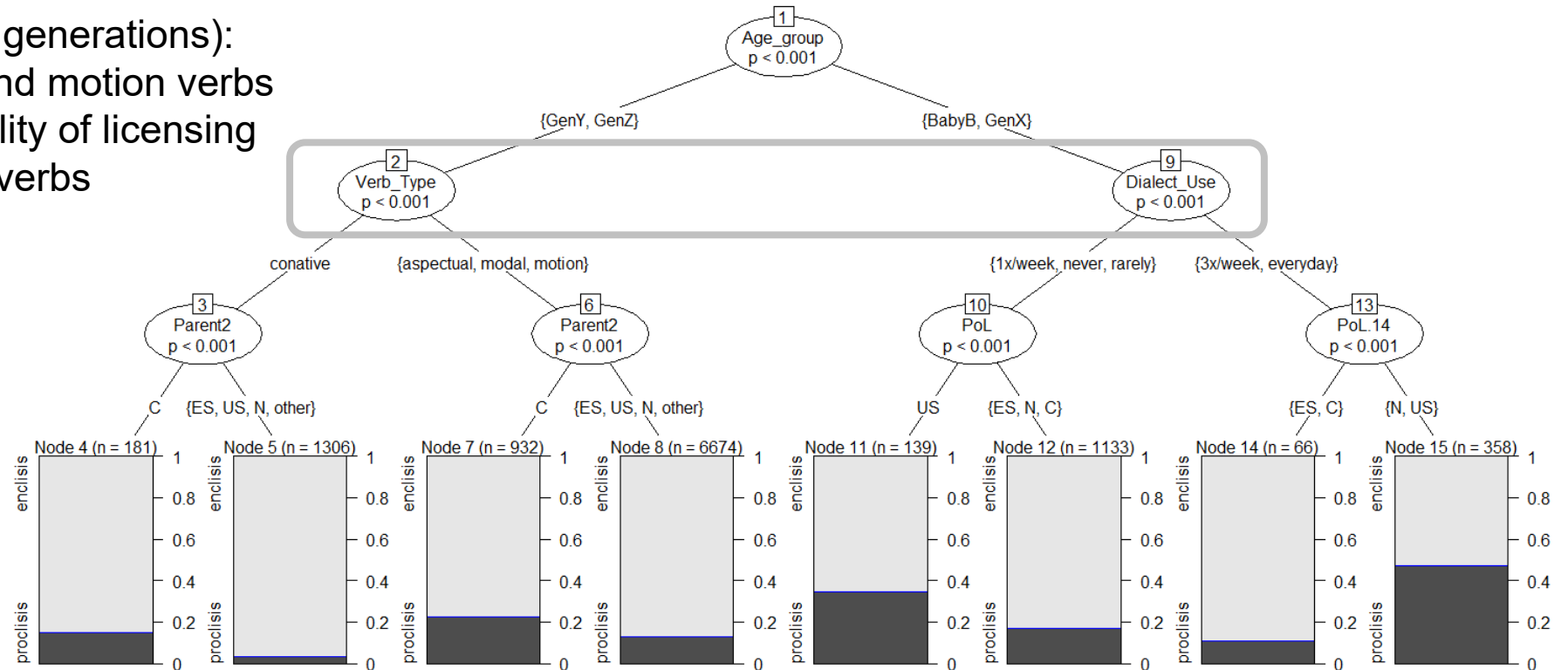
- age** is the most relevant predictor: older generations have higher probability of using OCC than younger generations



Data and Results

Conditional Inference Tree:

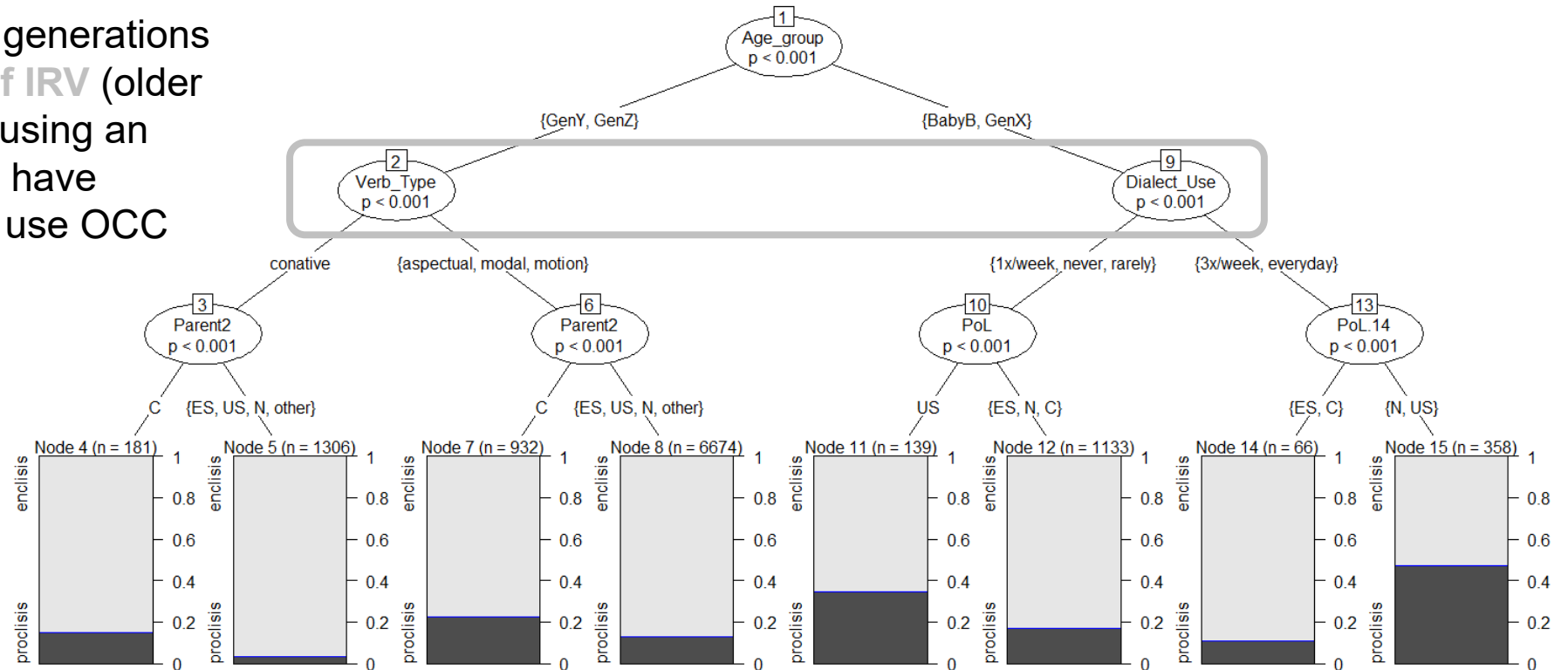
- **verb type** (younger generations):
aspectual, modal, and motion verbs
have higher probability of licensing
OCC than conative verbs



Data and Results

Conditional Inference Tree:

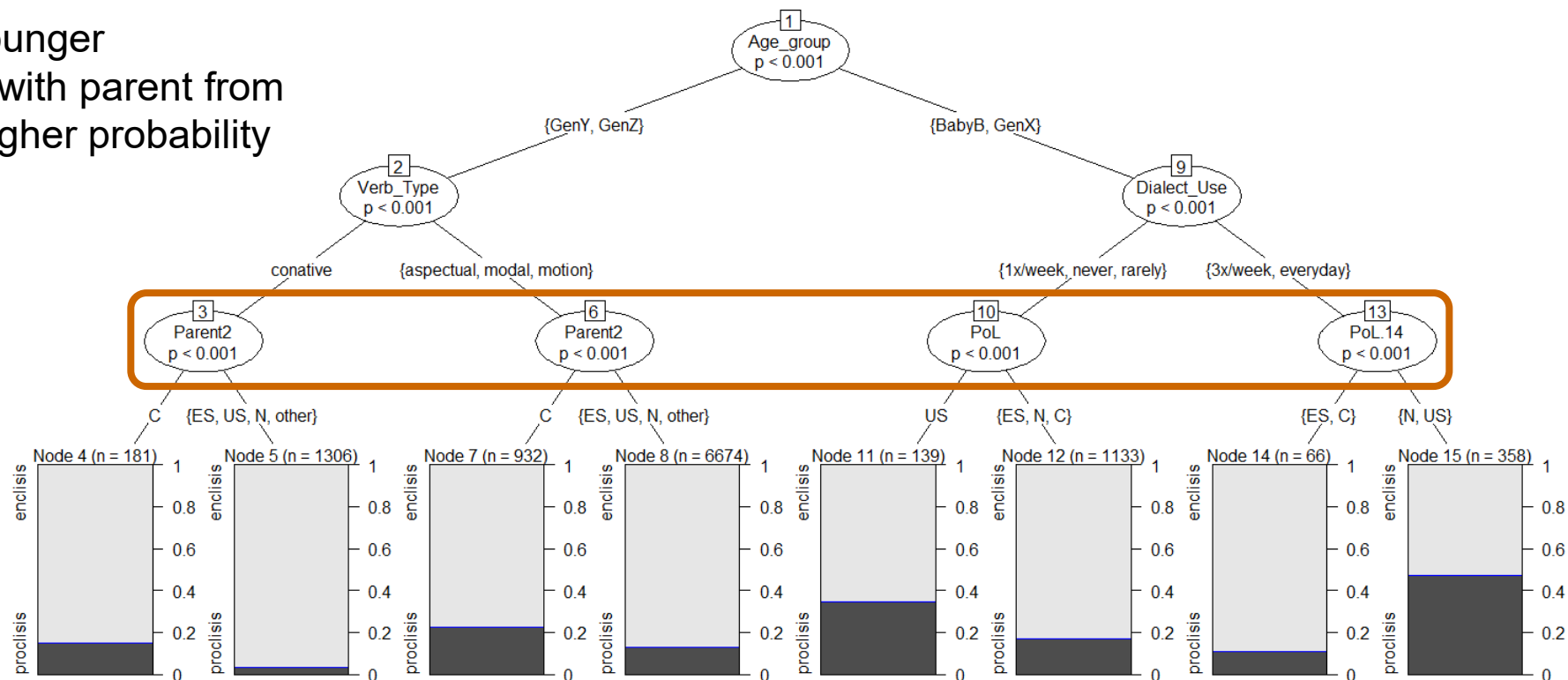
- **verb type** (younger generations)
- **frequency of use of IRV** (older generations): those using an IRV more frequently have higher probability to use OCC



Data and Results

Conditional Inference Tree:

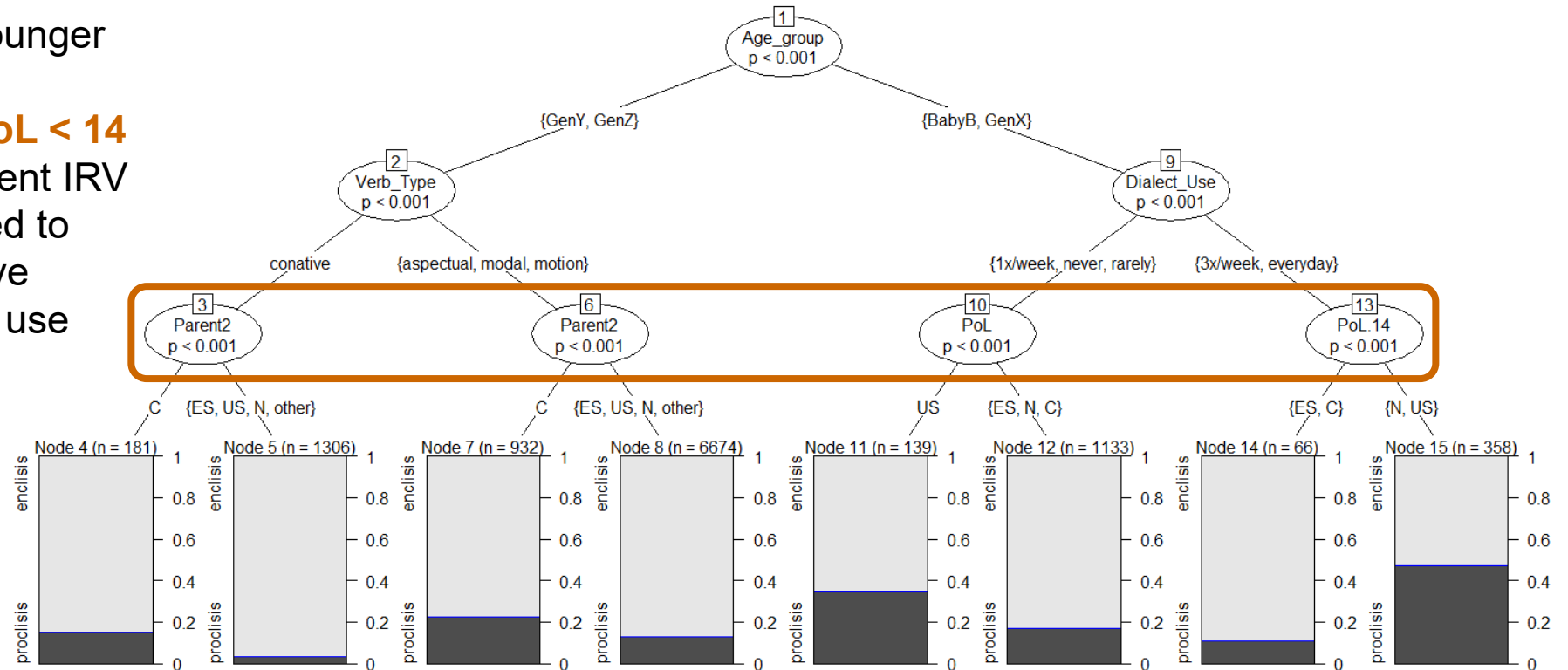
- origin of parent** (younger generations): those with parent from Central Italy have higher probability to use OCC



Data and Results

Conditional Inference Tree:

- **origin of parent** (younger generations)
- **place of living & PoL < 14** (older gens w. frequent IRV use): those who used to live in N and US have higher probability to use OCC

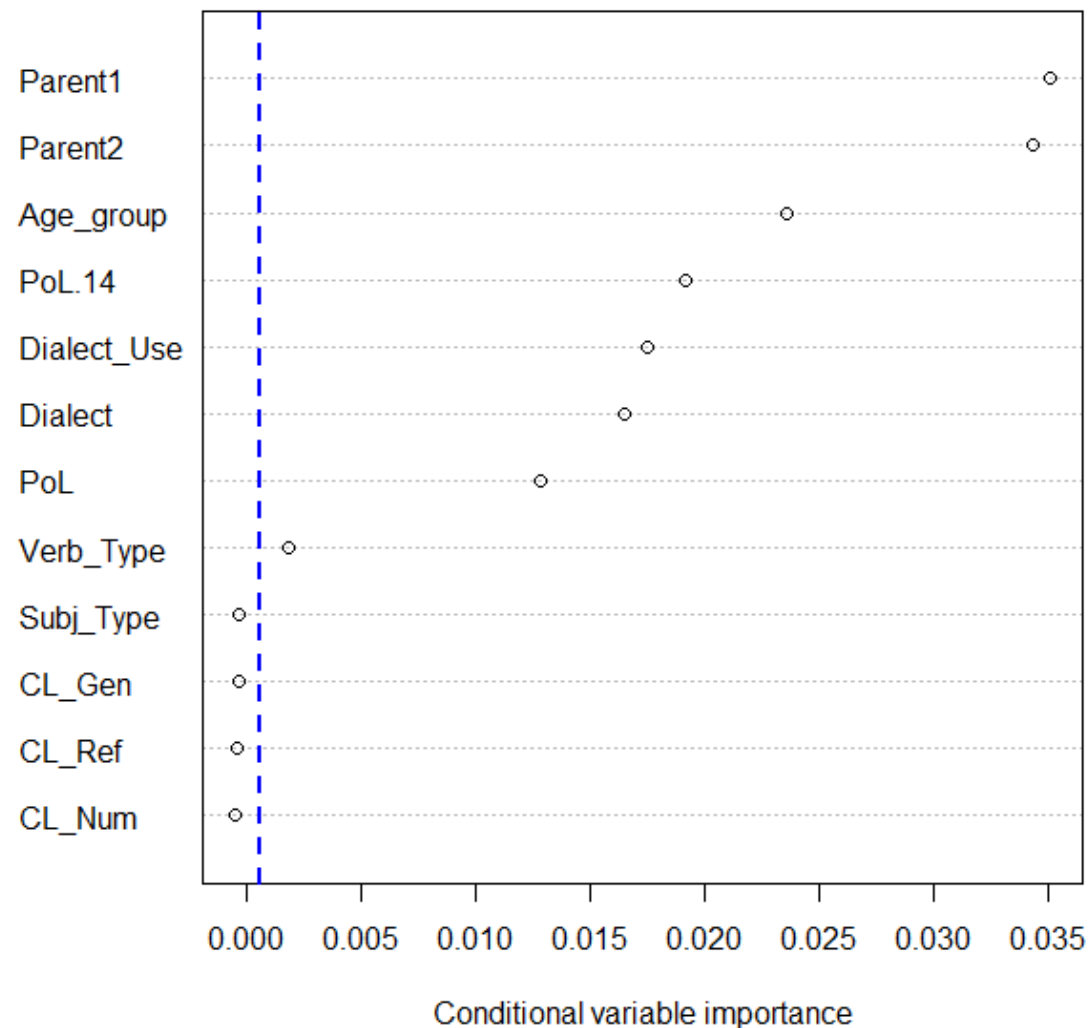




Data and Results

Conditional Random Forest:

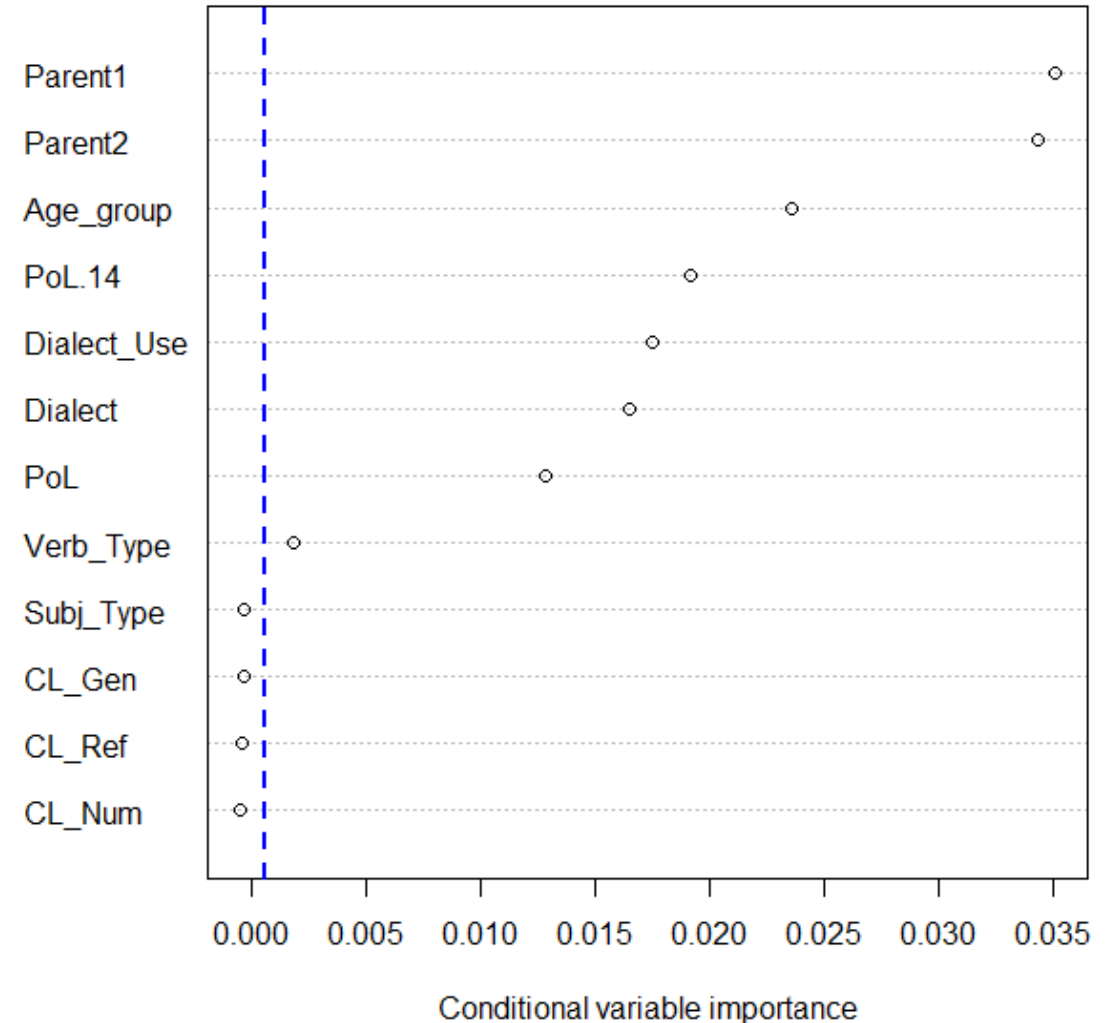
- forest of 2000 trees
- $C = 0.88$



Data and Results

Conditional Random Forest:

- linguistic variables are not predictive (exception: verb type)
- sociolinguistic variables are predictive:
 - linguistic background (incl. parents' origin and place of living before age 14) matter more than current place of living
 - age is relevant





Conclusion

- OCC is minority (proclisis 14% vs 81% enclisis)
- linguistic variables are not predictive (mild exception: verb type, in younger generations)
- sociolinguistic variables are predictive:
 - age is the most predictive variable
 - linguistic background (incl. parents' origin and place of living before age 14) is relevant
- *caveat*: these results are preliminary!



Conclusion

- OCC is minority (proclisis 14% vs 81% enclisis)
- linguistic variables are not predictive (mild exception: verb type, in younger generations)
- sociolinguistic variables are predictive:
 - age is the most predictive variable
 - linguistic background (incl. parents' origin and place of living before age 14) is relevant
- *caveat*: these results are preliminary—*watch this space, as we reach the 300-questionnaire threshold!*



References

Benincà, Paola. 1986. Punti di sintassi comparata dei dialetti italiani settentrionali. In G. Holtus & K. Ringger (eds), *Raetia antiqua et moderna. W Th.Elwert zum 80. Geburtstag*, 457-479. Tübingen: Niemeyer.

Calabrese, Andrea. 2001. I pronomi clitici. In Lorenzo Renzi, Giampaolo Salvi & Anna Cardinaletti (eds.), *Grande grammatica italiana di consultazione, vol. 1: La frase. I sintagmi nominale e preposizionale*, 563-606. Bologna: Il Mulino.

Cardinaletti, Anna. 2023. Variation and Optionality in Italo-Romance. Talk given at 17th CIDS, Zurich, Switzerland. URL: [https://www.spur.uzh.ch/en/veranstaltungen/workshops/2023/17th-Cambridge-Italian-Dialect-Syntax-Morphology-Meeting-\(CIDS-17\).html](https://www.spur.uzh.ch/en/veranstaltungen/workshops/2023/17th-Cambridge-Italian-Dialect-Syntax-Morphology-Meeting-(CIDS-17).html)

Cardinaletti, Anna & Giuliana Giusti. 2022. Dependency, licensing, and the nature of grammatical relations. In Adam Ledgeway & Martin Maiden (eds.), *The Cambridge Handbook of Romance Linguistics*, 604-636. Cambridge University Press.

Cardinaletti, Anna, Giuliana Giusti & Gianluca Lebari. 2022. Clitic climbing across Italy: Optionality as a function of bilingualism. Talk given at 16th CIDS, Naples, Italy. URL: <https://www.sfeezio.com/cids16/>

Cardinaletti, Anna & Uri Shlonsky. 2004. Clitic positions and restructuring in Italian. *Linguistic Inquiry* 35(4). 519-557.

Dragomirescu, Adina, Alexandru Nicolae & Gabriela P. Dindelegan. 2022. Complex predicates. In Adam Ledgeway & Martin Maiden (eds.), *The Cambridge Handbook of Romance Linguistics*, 570-603. Cambridge University Press.

Jones, Michael. 1993. *Sardinian syntax*. London: Routledge.



References

Ledgeway, Adam. 1998. La ristrutturazione in napoletano. In Giovanni Ruffino (ed.), *Atti del XXI Congresso Internazionale di Linguistica e Filologia Romanza*, 529-541. Tübingen: Niemeyer.

Ledgeway, Adam. 2000. A comparative syntax of the dialects of southern Italy: A minimalist approach. Oxford: Blackwell.

Manzini, Maria R. & Leonardo M. Savoia. 2005. *I dialetti italiani e romanci. Morfosintassi generativa: Volume 3: L'ausiliare (2a parte); la negazione e l'avverbio; strutture aspettuali e modali; il sintagma nominale; bibliografia*. Alessandria: Edizioni dell'Orso.

Monachesi, Paola. 1995. A grammar of Italian clitics. Tilburg: Universiteit Tilburg PhD dissertation.

Pescarini, Diego. 2021. *Romance object clitics*. Oxford: Oxford University Press.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria, <https://www.R-project.org/>.



Data and Results

CIT:

