Processing variable structures: 
Object clitics in native and L2 Spanish

Kimberly L. Geeslin & Tania Leal
Indiana University
Overarching goals

- Expand the body of interdisciplinary research connecting psycholinguistic methods to sociolinguistic inquiry to better understand:
  - The differences between native language processing and second language processing
  - Whether different variable structures in Spanish provide equally good test cases for processing effects
  - How geographically variable patterns influence real time sentence processing

- Move beyond existing strengths in the literature:
  - Focus on phonetics and English in social processing literature
  - Focus on production in research on Spanish
  - Treatment of “variable structure” as monolithic
  - Lack of focus on second language speakers and what the acquisition of sociolinguistic sensitivities can tell us about cognition more generally
Potential benefits

- Examination of how non-linguistic information guides comprehension:
  - Contributes to the growing understanding of whether social information is stored with linguistic information and how this influences processing (e.g., Casasanto 2008; Hay & Drager 2010; Squires 2013, 2014)

    - Early evidence that social and linguistic information influence perception; role for lexical and collocational frequency; talker influence

    - Studies show these factors matter, but the strength of influence, the directionality of influence and the nature of these relationships warrants further exploration
Potential benefits, continued

- Are variable structures processed in similar ways to categorical ones
  - Are unexpected forms associated with slower reading times? (in a parallel way to ungrammatical forms?)
  - How do we address allowable forms (experience) vs. forms a speaker would produce (preferences)? And how are these facts represented in mental models?
  - Does the social value (stigma, status of change in progress) influence how variable forms are processed?
  - Do geographic variants differ from social ones?
- If we are able to develop a method for testing the processing of variable structures that addresses these issues we can expand:
  - Role of Frequency: Are frequent combinations processed as quickly as infrequent ones when grammatical (linguistic) factors are held constant?
    (e.g. Bresnan & Ford, 2010: Processing of dative alternations in American and Australian English)
Clitic pronouns in Spanish

- In most varieties, 3rd person singular clitics share this paradigm

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
<th>Neuter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Masculine</td>
<td>Feminine</td>
<td>Masculine</td>
</tr>
<tr>
<td>Accusative</td>
<td>lo</td>
<td>la</td>
<td>los</td>
</tr>
<tr>
<td>Dative</td>
<td>le</td>
<td></td>
<td>les</td>
</tr>
</tbody>
</table>

- This **etymological clitic system** divides these clitics according to Case.

(e.g. Cuervo, 1988)
Clitic pronouns in Spanish

- A well-studied case of variation: leísmo
  - The (variable) use of dative pronoun le in accusative contexts
  - Spain: León, Castilla

Standard’ or Central’ Peninsular Spanish

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
<th>Neuter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masculine</td>
<td>Masculine</td>
<td>Masculine</td>
<td>Feminine</td>
</tr>
<tr>
<td>[+human]</td>
<td>[-human]</td>
<td>los</td>
<td>las</td>
</tr>
<tr>
<td>Accusative</td>
<td>le</td>
<td>la</td>
<td>lo</td>
</tr>
<tr>
<td>Dative</td>
<td>le</td>
<td>les</td>
<td>le</td>
</tr>
</tbody>
</table>

(Fernández-Ordóñez, 2012; Ormazabal & Romero, 2013)
Factors influencing leísmo

- **Telicity**
  - Positive relationship with telic verbs
    - Flores-Cervantes (2002); Geeslin et al. (2010); Parodi et al. (2012); Salgado-Robles (2014)

- **Animacy**
  - Positive relationship with animate referents
    - Geeslin et al., 2010; Salgado-Robles (2014)

- **Gender**
  - More acceptance with masculine referents

- **Co-referentiality**
  - Positive relationship in contexts of co-reference
    - Geeslin et al., 2010
The SLA of clitic pronouns

- **Successful acquisition of syntax (i.e., placement) of clitics**
  
  Bruhn de Garavito (2006); Bruhn de Garavito & Montrul (1996); Duffield & White (1999); Liceras (1985)

- **Clitic pronoun morphology lags behind**
  
  Franceschina (2001); McCarthy (2002)

- **Learner *defaults* or *prototypes* initially**
  
  Andersen (1984); Klee (1989); Malovrh (2008); Malovrh & Lee (2009); Lee & Malovrh (2009); Zyzik (2006)
The SLA of clitic pronouns

- Accuracy rates may reflect different hierarchies of features in development
  - 1p/2p > 3p
    - Andersen (1984); Klee (1989); Malovrh & Lee (2010)
  - Dative > Accusative
    - VanPatten (1990); Malovrh & Lee (2010)
  - Masculine > Feminine
    - Andersen (1984); Klee (1989); Malovrh (2008); Malovrh & Lee (2010)
  - Case > Gender
    - Franceschina (2001); Klee (1989); Malovrh & Lee (2010)
  - Person > Number > Gender (Carminati, 2005; Greenberg, 1963)
    - Malovrh & Lee (2010)
  - Underpsecification of Gender & Number (Harley & Ritter, 2002)
    - McCarthy (2008)
The SLA of clitic pronouns

**Zyzik (2006):** Overgeneralization of *le* in accusative contexts
- Intermediate: (22%); High intermediate (37.6%): Advanced (46.2%)
- Use of *le* in animate contexts reached over 90%
- Latin American or no SA (study abroad) experience

**Malovrh (2008):** Development of accusative pronouns
- Dative pronouns in accusative contexts
- Unable to connect this to language background; low number of tokens

The upshot: *le* in accusative contexts is attested, especially with animates, with no clear role for geography

See Malovrh (2013) for a detailed review
The SLA of clitic pronouns: Study abroad

Geeslin et al. 2010:

- 8-week intensive SA experience in León, Spain
- Contextualized preference task
- Frequency of *le*:
  - L2: 58.6% (Time 1), 41.4% (Time 2), 46.6% (Time 3)
  - NS: 54.4%

Predictors:

(adapted for space)

<table>
<thead>
<tr>
<th></th>
<th>NNS Time 1</th>
<th>NNS Time 2</th>
<th>NNS Time 3</th>
<th>NS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td>X***</td>
</tr>
<tr>
<td>Co-Ref</td>
<td>X***</td>
<td></td>
<td>X***</td>
<td></td>
</tr>
<tr>
<td>Subj. Animacy</td>
<td></td>
<td></td>
<td>X***</td>
<td>X***</td>
</tr>
<tr>
<td>Telicity</td>
<td>X***</td>
<td>X***</td>
<td>X*</td>
<td>X***</td>
</tr>
<tr>
<td>Improvement</td>
<td>X*</td>
<td></td>
<td>X***</td>
<td></td>
</tr>
</tbody>
</table>

Note. *** = p ≤ 0.001; ** = p ≤ 0.001; * = p ≤ 0.005
The SLA of clitic pronouns: Study abroad

**Salgado-Robles (2011): Two contexts**

- SA learners in Valladolid & Seville, Spain
- NS comparison groups at each site
  - Valladolid = Leísta system
  - Seville = Etymological/non-leísta system
- Pre- & post-interview data (5 months)
- Frequency
  - Valladolid: pre = 17.9% / post = 41.5% (NSs = 62.6%)
  - Seville: Pre = 16.4% / post = 12.4% (NSs = 10.9%)
- Predictors:
  - Referent gender for Valladolid (not Seville)
How can processing research tell us more?

- Sentence- and speech processing research showing that both linguistic and non-linguistic information shapes expectations. (e.g. Tannenhaus et al., 1995; Squires, 2014)

- Casasanto (2008): Overall reaction times
  - Nonlinguistic context (black vs. white speakers in t/d deletion),
  - Evidence of quick integration of this knowledge which helped speakers resolve ambiguity.

- Squires (2013)
  - Speaker’s perceptions of patterns of agreement were influenced by linguistic but not social primes
  - “Limited bidirectionality” between linguistic variables and social features.
Methodology: Self-paced reading

- Participants read sentences in a word-by-word or segment-by-segment presentation.
- To move along, they press a button/key (hence self-paced).
  - Participants can control the rate of presentation but cannot return to the previous segment/word.
- In general, higher (i.e. slower) reading times are believed to represent some sort of processing difficulty:
  - ungrammaticality
  - violation of an expectation
  - reanalysis processes

(Just, Carpenter & Wooley, 1982).
Processing research and second language acquisition of variable structures

- Together, previous research shows that the psycholinguistic methods can successfully reveal information about language processing.

- However, the methods that we use to examine linguistic variables may or may not be sensitive enough (or authentic enough) to capture the influence of social or geographic factors.

- **Our research:**
  
  *Explore whether the self-paced reading paradigm can contribute to our understanding of the influence of social and linguistic factors on a geographically-variable structure in Spanish.*
### Phases of this project

<table>
<thead>
<tr>
<th>Participants</th>
<th>Phase 1 (PILOT)</th>
<th>Phase 2</th>
<th>Phase 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>US bilinguals</td>
<td>US bilinguals (L1 Spanish, many varieties)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>L2 learners – 3 levels</td>
<td>(No learners)</td>
<td>Mexico, Seville (Spain)</td>
<td>Mexico, León (Spain)</td>
</tr>
<tr>
<td>Stimuli</td>
<td>Antecedents are self-paced</td>
<td>Antecedents in the context</td>
<td>Antecedents in close proximity (and self paced)</td>
</tr>
<tr>
<td></td>
<td>Few items (pilot)</td>
<td>Few items</td>
<td>More items needed</td>
</tr>
</tbody>
</table>
Phase 1: Research Questions

1. When manipulating two variables, pronominal form and gender, do self-paced reading times vary from one condition (of 4) to another?
   - Do these differ by linguistic condition?
   - Do these differ by participant group (NS vs. L2)?

2. Can second language learners acquire these same sensitivities?
Phase 1: Methodology

- **Participants**
  - Native speakers of Spanish (Span-Eng bilinguals)
  - Educated young adult readers from leista and non-leista regions
  - Second language learners (comparable age & SEC)

- **Tasks**
  - Level test (0-25 points)
  - Background questionnaire
  - Self-paced reading task
    - Conditions (2x2):
      - gender (masc./fem.)
      - pronoun (lo/la).
    - Factors controlled: Animate, 3rd p. sg., co-referential, telic; use of reading removes additional acoustic cues that might influence processing
Phase 1: Experimental manipulation

Context: My friends and I want to have dinner at a small restaurant and we are deciding who should come with us. One of them suggests:

Invitemos a Laura. / Es que / rara vez / la visito / porque / vive lejos / de la ciudad.
Invitemos a Martín. / Es que / rara vez / lo visito / porque / vive lejos / de la ciudad.
Invitemos a Laura. / Es que / rara vez / le visito / porque / vive lejos / de la ciudad.
Invitemos a Martín. / Es que / rara vez / le visito / porque / vive lejos / de la ciudad.

“It’s because I rarely visit him/her since he/she (pro) lives far from the city.”
Phase 1: Predictions

- **Natives**: Processing variants outside of one’s own increases difficulty. Thus, we expect
  1. non-\textit{leísta} speakers will read “\textit{le}” clitics more slowly than “\textit{lo/la}” clitics;
  2. gender effects: dative clitics + fem. referents will elicit slower RTs.

- **L2ers**: previous work shows:
  1. preference for masculine forms (across contexts)
  2. gender not a significant predictor of le in accusative contexts
  3. “\textit{le}” clitics interpreted as simply [+human].

Thus, we predict that all learners may process masculine and/or “\textit{le}” forms faster; at higher levels, feminine referents may be processed more slowly according to experience.
Table 1: Predictions for RTs (region of interest) in all conditions: natives and learners

<table>
<thead>
<tr>
<th></th>
<th>NS</th>
<th>L2ers (lower prof.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: &quot;la&quot;+fem ref.</td>
<td>Low RTs</td>
<td>Highest RTs</td>
</tr>
<tr>
<td>B: &quot;lo&quot;+masc. ref.</td>
<td>Lowest RTs</td>
<td>High RTS</td>
</tr>
<tr>
<td>C: &quot;le&quot;+fem ref.</td>
<td>Highest RTs</td>
<td>Low RTs</td>
</tr>
<tr>
<td>D: &quot;le&quot;+masc. ref</td>
<td>High RTs (depending on variety)</td>
<td>Lower RTs</td>
</tr>
</tbody>
</table>
Phase 1: Results

2-way RM ANOVA – within-participant vbles pronoun (“lo”/ “le”) gender (m/f)
- between-participants vble Group
Effects and interactions
Gender (F(1, 60)=5.808, p<0.019)
three-way interaction (pronoun*gender*group; F(3,60)=2.794, p=.048).
Phase 1: Results

- **Natives** (U.S. bilinguals from various countries):
  1. Marginally read “le” forms more slowly than “la” forms (with feminine referents) \((p=0.081)\) → expected
  2. but not with masculine referents \((p=.461)\)

- **Intermediate-low learners:**
  1. Read “le” with feminine referents more quickly than “la” forms (fem. referents; \(p=0.030\)). *Opposite pattern from native speakers.*

- **Intermediate learners:**
  1. Read “le” with masculine referents more quickly than “lo” forms (masc. referents; \(p=0.003\)) → mirroring the developmental literature

- **Advanced learners:**
  1. Marginally read “le” forms more slowly only with masculine referents (unlike natives but like intermediates) \((p=0.09)\) but not feminine referents \((p=.382)\)
Phase 1: Interpretation

- These results suggest that L2 learners display sensitivity to only some linguistic factors that influence NS-choices in variable structures.

- The methodology seemed to work for this particular structure (not others)

- Too few items (Latin square presentation)
Phase 2: Motivation and Issues w/Phase 1

- Presentation of the items worked best for le/lo forms
  - We moved the antecedents from the self-paced (close) section to the context to alleviate issues with length and frequency for other structures

- Our native speakers were from several countries of origin, making it difficult to tease apart the influence of linguistic, social and geographic factors
  - Phase 2 = focus on single-variety native-speaker groups
Phase 2: Research Questions

1. When manipulating two variables, pronominal form and gender, do self-paced reading times vary from one condition (of 4) to another?
   ➡️ Do these differ by linguistic condition?
   ➡️ Do these differ by participant group (Mexico vs. Seville)?
Phase 2: NSs and task modifications

- **Participants**
  - Native single-variety speakers of Spanish
    - Mexico
    - Spain (Seville)

- **Tasks**
  - Level test
  - Background questionnaire
  - Self-paced reading task (multiple structures)
    - Conditions (2x2):
      - gender (masc./fem.)
      - form (“lo”/“le”).
    - Antecedent in CONTEXT
Phase 2: Experimental manipulation

Context: My friends and I want to have dinner at a small restaurant and we are deciding who should come with us. One of them suggests inviting Laura/Martín, saying the following:

“Es que / rara vez / **lo visito** / porque / vive lejos / de la ciudad.”

“Es que / rara vez / **la visita** / porque / vive lejos / de la ciudad.”

“Es que / rara vez / **le visito** / porque / vive lejos / de la ciudad.”

“Es que / rara vez / **le visito** / porque / vive lejos / de la ciudad.”

“It’s because I rarely visit him/her since he/she (pro) lives far from the city.”
Predictions for RTs (region of interest) in all conditions: natives (Mexico and Seville)

<table>
<thead>
<tr>
<th></th>
<th>Mexico</th>
<th>Seville</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: &quot;la&quot;+fem ref.</td>
<td>Low RTs</td>
<td>Low RTs</td>
</tr>
<tr>
<td>B: &quot;lo&quot;+masc. ref.</td>
<td>Lowest RTs</td>
<td>Lower RTs</td>
</tr>
<tr>
<td>C: &quot;le&quot;+fem ref.</td>
<td>Highest RTs</td>
<td>Low RTs</td>
</tr>
<tr>
<td>D: &quot;le&quot;+masc. ref</td>
<td>High RTs (depending on variety)</td>
<td>Lower RTs</td>
</tr>
</tbody>
</table>
Phase 2: Results

2-way RM ANOVA
- within-participant vbles
  pronoun (“lo”/”le”)
  gender (m/f)
- between-participants vble
  Group

Effects
Group
F(1, 54)=9.286, p<0.01

Interactions
Gender*Group
F(1, 54)=6.039, p=0.017

Pronoun*Group;
F(1,54)=4.577, p=.037).

RESIDUALS = (RAW times) – (PREDICTED times)
- Positive: slower than predicted
- Negative: faster than predicted
Phase 2: Results

- **Mexican NS**
  1. Read “le” clitics more slowly,
     (pairwise Bonferroni comparisons $p=0.055$)
  2. But there was no gender distinction
     (pairwise Bonferroni comparisons $p=0.203$)

- **Conversely, Sevillians**
  1. Marginally read “lo” and “le” pronouns differently → “le” forms were faster, unlike Mexicans.
     (pairwise Bonferroni comparisons $p=0.055$)
  2. But did make a robust distinction between feminine and masculine referents, reading masculine faster overall.
     (pairwise Bonferroni comparisons $p=.024$)
Discussion: Research questions

1. When manipulating two variables, form and gender, do self-paced reading times vary from one condition (of 4) to another

- Do these differ by linguistic condition?
  
  YES but effects varied a bit. With improved materials, we believe we have uncovered an area that is ripe for investigation

- Do they differ by geographic background/experience of the speaker?

  YES but our pilots highlight the need to get more background information and to perform norming studies with the totality of the tokens
Discussion: Research Questions

2. Can second language learners acquire these same sensitivities?

- Not in Phase 1. The advanced learners show a different pattern from NS but there is evidence of development over time. More systematic background information will help us to answer this question.
Methodological Challenges

- Too few tokens: Could be an issue with replicability
- Verbs (lexical items) need to be checked for frequency of occurrence with the "le" clitic form
- Mexican speakers were bilingual (Spanish-French) which could create a confound (French clitics function in a different way—"le" is masculine accusative but "la" overlaps)
- Distance of the referent must be further explored (effects varied slightly from Phase 1 to 2 [varying degrees for different structures])
Phase 3 (planning stage): a return to background information

- With the knowledge that...
  - the placement of the antecedent might matter for the predictions for clitic form alternations
  - we did not test exclusively leísta regions

- Additional goal:
  - Examine learners at the highest level who have at least 6 months experience with native speakers from a leísta region and those who have similar experience with a non-leísta variety in order to better understand the group behavior from the Phase 1 results

- Double items after limiting range of structures
## Phases of this project

<table>
<thead>
<tr>
<th>Phase 1 (PILOT)</th>
<th>Phase 2</th>
<th>Phase 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participants</strong></td>
<td><strong>Stimuli</strong></td>
<td><strong>Challenges</strong></td>
</tr>
<tr>
<td>US bilinguals</td>
<td>Antecedents are self-paced</td>
<td>Number of items (too few)</td>
</tr>
<tr>
<td>• (L1 Spanish, many varieties)</td>
<td>• Antecedents in the context</td>
<td>• Other structures</td>
</tr>
<tr>
<td>L2 learners – 3 levels</td>
<td></td>
<td>• Interpreting results</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Number of items (too few)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Interference from French</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• No “pure” leísta region tested</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS</td>
<td></td>
<td>Antecedents in close proximity (and self paced)</td>
</tr>
<tr>
<td>• Mexico,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Seville (Spain)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(No learners)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Mexico,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• León (Spain)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L2 learners in US</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Continuing the narrative: a tale of cross-disciplinary tensions and promise

<table>
<thead>
<tr>
<th></th>
<th>Sociolinguistic variation</th>
<th>Self-paced reading (or other similar measures)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Assumptions</strong></td>
<td>Sociolinguistic variation does not come at a cost (although unexpected variants might)</td>
<td>Unexpected forms lead to slower reading times</td>
</tr>
<tr>
<td><strong>Goals</strong></td>
<td>Understand how sociolinguistically-variable forms are processed</td>
<td>Determine the place for sociolinguistic information in models of sentence processing (vs. information encapsulation)</td>
</tr>
<tr>
<td><strong>Methodological requirements</strong></td>
<td>Unmonitored production data vs. implicit attitudinal methods context-driven interpretation</td>
<td>Implicit measures of tightly manipulated stimuli in several linguistic dimensions</td>
</tr>
</tbody>
</table>
Thank you!

- To the graduate students for support of L2 data collection on campus
- To field researchers Angel Milla and Fernando Melero Garcia
- Research assistants Jordan Garrett and Travis Sago
- This research was supported by an Indiana University SEED grant
Selected References


