Pronominal subjects at the interfaces: evidence from American English learners of Spanish in the CEDEL2 corpus.

Cristóbal Lozano
Universidad de Granada
http://www.ugr.es/~cristoballozano

ISB6 (6th International Symposium on Bilingualism)
30 May-2 June 2007 Hamburg, Germany

THE PHENOMENON:
Pronominal subjects in L1 English - L2 Spanish

- Early findings (1980s)
  - Licensing mechanisms: acquired early
    - Learners of L2 Spanish know from early stages that overt and null pronouns can alternate.

- Recent findings (late 1990s, 2000s):
  - Discursive properties acquired late and persistently problematic
  - Distribution of overt and null pronouns constrained by discourse
  - Deficits at syntax-discourse interface

Previous studies: SLA

- Formal features: Learners omit (=license) pronominal subjects from early stages of acquisition
    "El/él come pasta" or "He/He eats pasta"
  - Discursive features: Learners acquire the discursive constraints (topic, topic-shift) late: persistently problematic → overproduction in topic contexts
      - "La abuela piensa que ella regará las plantas"
      - "Grandma thinks that she will water the plants"

- Unidirectionality:
  - Most studies have found overproduction of overt subjects in topic contexts
  - But some bidirectionality attested as well (Montrul & Rodriguez-Louro 2006)

Previous studies: Attrition

- e.g., Toribio 2000, Satterfield 2003
  - Overproduction of overt subjects in topic contexts:
    "Yo no lo vi, él estaba en Massachusetts ... él estaba en New York ... él estaba donde Eugenia ..."
    - "I didn't see him, he was in Massachusetts ... he was in New York ... He was at Eugenia's place ..."

  Quando ellos vienen aquí, ellos lo pierden
  - "When they arrive here, they lose it (i.e., they lose their mother tongue Spanish)"
The wider context: Syntax before discourse

- The “syntax-before-discourse” phenomenon
  - Observed in 2 properties of pro-drop parameter
- Observed in other acquisition contexts:
  - L2 acquisition of pronominals (other than L2 Spa):
  - L2 acquisition of SV inversion
  - L1 bilingualism of pronominals
  - L1 attrition of pronominals and SV inversion:
  - L1 acquisition of pronominals

- See Sorace (2004) for overview:
  - Interfaces, because they are more complex than narrow syntax, are inherently more difficult to acquire” (p. 144)

Explaining the causes

  - 2 main approaches:
      - Underspecification of (interpretable) features like [Topic], [Focus], [Topic-Shift]
    - PROCESSING DEFICITS: Processing deficits at syntax-discourse interface (L2 Spa: Lozano 2006c)
      - Default processing mechanism: overt pronoun

STILL MISSING IN THE LITERATURE

- In previous studies:
  - Most evidence comes from 3rd singular only
  - BUT claims about the whole pronominal paradigm (singular: 1st, 2nd, 3rd; plural: 1st, 2nd, 3rd)
- This study:
  - Evidence from the whole paradigm
  - Deficits are selective:
    - Not all persons in the paradigm are equally vulnerable at syntax-discourse interface
  - Birectinality:
    - Both overt and null pronouns are used apragmatically.

FEATURE GEOMETRY ANALYSIS
Harley & Ritter (2002)

- Pronominal feature geometry: part of UG ➔ hierarchically organised

- Traditional distinction (Jespersen, Bühler, Bloomfield, Benveniste, etc):
  - Speech-act participants vs. non-participants
    - Speech-act participants: Deictic use of pronouns (1st, 2nd)
    - Speech act non-participants: Anaphoric use of pronouns (3rd)
PRONOMINAL DISTRIBUTION at SYNTAX-DISCOURSE:

TOPIC
- Topic = known information
- Topic continuity is marked via null pronoun

[Previous context: talking about the main character of a film]

En la película “Escondido” el protagonista tiene una familia y trabaja en un programa de televisión. Un día empieza a recibir videos anónimos...

Theoretical basis for pronominal distrib in discourse:

DATA:
CEDEL2 corpus
[Corpus Escrito del Español L2]
- L1 English – L2 Spanish
- Also: native Spanish subcorpus (for comparative purposes)
- All proficiency levels
- 3 online forms:
  - Learning background
  - Spanish placement test (UWPT: University of Wisconsin Placement Test)
  - Composition

http://www.uam.es/woslac/start.htm
DATA:
CEDEL2 corpus:

Number of words

Mar-06
Apr-06
May-06
Jun-06
Jul-06
Aug-06
Sep-06
Oct-06
Nov-06
Dec-06
Jan-07
Feb-07
Mar-07
Apr-07
May-07

SOFWARE:
UAM Corpus Tool, v. 1.0

FREE SOFTWARE: http://www.wagsoft.com/CorpusTool/

- Developed by Mick O’Donnell (Universidad Autonoma de Madrid)
- Annotates segments
  - A segment can be any linguistic unit:
    - Morpheme
    - Word
    - Phrase
    - Sentence
    - Text
  - Each annotation is a tag.
  - XML-compatible.

ANNOTATION SCHEME (tags)

SOFTWARE: TAGGING
HYPOTHESES

- 2 HYPOTHESES
  - Based on previous research and the Feature Geometry:
    - H1: Deficits at syntax-discourse interface are selective:
      - Robustness and native-like knowledge:
        - speech-act participants: 1st and 2nd
      - 3rd neutral
      - Vulnerability: 3rd animate
  - H2: Bidirectionality
    - Overuse of overt pronouns in topic contexts
    - Underuse of null pronouns if topic-shift contexts

SUBJECTS

- 2 learner groups (L1 AmEng – L2 Spa): Students of L2 Spa at American Universities, placement program in Spa-speaking country.
  - Upper Advanced: 98%-100% proficiency
  - Lower Advanced: 91%-95% proficiency
- 1 Spanish native control group
- Proficiency measure: UWPT (University of Wisconsin Placement Test, 1998)

Written compositions:
- For this study, mostly:
  1. Talk about one of your recent travels. [1st person]
  2. Summarise a film you have seen recently. [Topic, Topic-Shift, 3rd person]
  3. Talk about your last holidays. [1st person, 3rd person, Topic, Topic-shift]

RESULTS

DATA ANALYSIS: analysis of frequencies χ² inferential stats to check significance
RESULTS: 1sing Animate

yo/Ø ‘I’

- All groups: Animate
- All groups: mostly pragmatically correct
- No sig differences between groups

CONCLUSION:
- Low-Adv = Spa
- Upp-Adv = Spa

RESULTS: 1plu Animate

nosotros/nosotras/Ø ‘we’

- All groups: Animate
- All groups: pragmatically correct
- NO Sig differences between learners and Spanish

CONCLUSION:
- Low-Adv = Spa
- Upp-Adv = Spa

RESULTS: 2sing, 2plu

tú/Ø ‘you’
vosotros/vosotras/Ø ‘you all’

- 2nd sing: tú ‘you’
  - All groups: Animate
  - All groups: pragmatically correct
  - All groups: Extremely low freq, though:
    - Upp-Adv: N=1
    - Low-Adv: N=1
    - Spanish: N=0

- 2nd plu: vosotros / vosotras ‘you all’
  - All groups: Animate
  - All groups: pragmatically correct
  - All groups: Extremely low freq, though:
    - Upp-Adv: N=1
    - Low-Adv: N=1
    - Spanish: N=0

RESULTS: 3sing Animate

NP/él/ella/Ø ‘he/she’

- Learners: considerable % of pragmatically incorrect (17%, 15%)
- Natives: pragmatically correct
- Sig differences between learners and Spanish

CONCLUSION:
- Low-Adv ≠ Spa
- Upp-Adv ≠ Spa
RESULTS: 3sing Inanimate
NP/ello/Ø ‘it’

- All groups: pragmatically correct
- NO sig differences between learners and Spanish.

CONCLUSION:
- Low-Adv = Spa
- Upp-Adv = Spa

RESULTS: 3plu Animate
NP/ellos/ellas/Ø ‘they’

- All groups: mostly pragmatically correct
- NO sig differences between learners and Spanish.
- BUT: learners higher apragmatic than Spanish (10%, 10%)

CONCLUSION:
- Low-Adv = Spa
- Upp-Adv = Spa

RESULTS: 3plu INANIMATE
NP/Ø ‘they’

- All groups: pragmatically correct (except for 4.5% in Low-Adv, which corresponds to 1 example only)
- NO sig differences between learners and Spanish.

CONCLUSION:
- Low-Adv = Spa
- Upp-Adv = Spa

SUMMARY OF RESULTS:
PERSON & NUMBER

<table>
<thead>
<tr>
<th></th>
<th>SINGULAR</th>
<th>PLURAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>2nd</td>
<td>No problem. Learners behave statistically like Spanish. [BUT more research needed due to low frequencies]</td>
<td>No problem. Learners behave statistically like Spanish. [BUT more research needed due to low frequencies]</td>
</tr>
<tr>
<td>3rd</td>
<td>Problematic: Animate Learners behave differently from Spanish.</td>
<td>No problem: Inanimate Learners behave statistically like Spanish.</td>
</tr>
</tbody>
</table>
RESULTS: Apragmaticality of Topic

- Overproduction of overt pronoun in topic-continuity context:
  - [Context: Talking about two characters in the film "Volver": the mother and Penélope Cruz]
    La madre ha vuelto para pedirle perdón a Penélope, porque ella no sabía que Penélope había sido abusada por su propio padre... [JEL, Upp-Advanced, CEDEL2 corpus]
  - [Context: Talking about the main character of the film "Spanglish"]
    La madre no puede hablar inglés pero ella es muy trabajadora. #Ella empieza a trabajar ...
    #Ella no puede comunicar(se) con esta familia..... [SMM, Low-Advanced, CEDEL2 corpus]

- Overproduction of full NP in topic-continuity context:
  - [Context: Talking about the main character of the film "Spanglish"]
    La trabajadora está muy celosa y no le gusta la situación. A pesar de la situación, #la trabajadora se queda en la casa y Ø continúa a trabajar...
    #Ella no puede comunicar(se) con esta familia..... [SMM, Low-Advanced, CEDEL2 corpus]

RESULTS: Apragmaticality of Topic-Shift

- Underproduction of overt pronoun in topic-shift contexts (=production of Ø) ➔ ambiguity
  - [Context: Describing the film "Amores posibles": they are various characters: Carlos, Julia, María, la madre (de Julia), el padre (de Carlos)]
    En la tercera [película], Ø deja su romance con Julia y Ø regresa a su esposa, que en mi opinión Ø sabía lo que estaba pasando con su esposo pero nunca Ø quiso decirle nada sabiendo que Ø tenía [que] hacer su propia decisión de volver a estar con ella.

CONCLUSION
- To mark topic continuity, learners significantly overproduce overt material when a null pronoun is required.
RESULTS: Apragmaticality of Topic-Shift

- Learners: very low percentages of underproduction (3%, 8%)
- Natives: no underproduction (0%)
- Learners do not statistically differ from natives:
  - Up-Adv = Spanish
  - Low-Adv = Spanish
- BUT learners do commit some underproduction violations.

CONCLUSION

- Technically, learners behave like Spa natives, though they produce some underproductions.

CONCLUSION

- Previous research:
  - Deficits when overt/null pronouns are constrained by topic and topic-shift at syntax-discourse interface
- Implication: Deficits with Economy Principle: no underproduction
- Chomsky’s ‘Avoid Pronoun Principle’
- Deficits at syntax-discourse interface are SELECTIVE:
  - Robust 1st and 2nd persons (sing & plu)
  - speech-act participants
  - deictic use of pronouns
  - Robust 3rd person inanimate (sing)
  - Vulnerable 3rd person animate (sing & plu)
  - Anaphoric use of pronouns

Further research:

- Pronominal Feature Geometry (Harley & Ritter 2002) and SLA of pronouns
- Limitations of corpus data. Experimental data needed (whole pronoun paradigm)