

Pronominal deficits at the interface: new data from the CEDEL2 corpus

CRISTÓBAL LOZANO

Universidad de Granada

ABSTRACT

Recent studies reveal that learners of L2 Spanish are sensitive to the formal syntactic mechanisms licensing overt and null pronominal subjects from early stages of acquisition, but they show residual deficits when their distribution is constrained by topic and focus at the syntax-discourse interface, even at advanced levels of proficiency. Importantly, previous research has assumed that all phi-features of the pronominal paradigm are equally vulnerable, but the current paper presents data from CEDEL2 showing that deficits are selective as they affect 3rd person animate features only.

RESUMEN

Según recientes estudios los aprendices de español L2 muestran déficits en la interfaz sintaxis-discurso cuando la distribución pronombres sujeto plenos y nulos está regulada por el discurso, aunque adquieren desde etapas tempranas los mecanismos formales (sintácticos) que regulan dicha distribución. Los estudios previos han asumido que todos los pronombres del paradigma son propensos a estos déficits, pero aquí mostramos, usando datos de CEDEL2, que los déficits son selectivos ya que afectan sólo a los rasgos 3^a persona masculino/femenino.

KEYWORDS: *learner corpora, CEDEL2 corpus, Spanish L2 acquisition, syntax-discourse interface, pro-drop, null subjects*

1 INTRODUCTION

English-speaking learners of L2 Spanish acquire early the *formal* features that license overt/null referential pronominal subjects, but even at advanced and end-state they show deficits with the *discursive* features that license their distribution in the discourse (see section 3).

I will show that the observed deficits are selective, since only 3rd person singular animate (masculine and feminine) is vulnerable, while the rest of persons (even 3rd singular neutral) remain intact.

2 THEORETICAL BACKGROUND

2.1 *Distribution of pronominal subjects at the syntax-discourse interface*

The apparently free alternation of overt and null pronominal subjects in pro-drop languages like Spanish, (1a), is constrained discursively (Alonso-Ovalle et al. 2002, Fernández-Soriano 1989). A null pronoun (*pro*), (2), encodes [Topic-continuity]: References to the already-mentioned topic *el protagonista* are realized with a null pronoun.

- (1) a. $\left. \begin{array}{c} \text{Pedro} \\ \text{él} \\ \text{pro} \end{array} \right\}$ tiene mucho dinero.
- b. Peter/he/* \emptyset have.3sg a lot of money
'Pedro/he has a lot of money'
- (2) En la película "Escondido" **el protagonista** tiene una familia y [**pro**] trabaja en un programa de televisión. Un día [**pro**] empieza a recibir videos anónimos [...] [RSZ, Spanish native, CEDEL2 corpus]

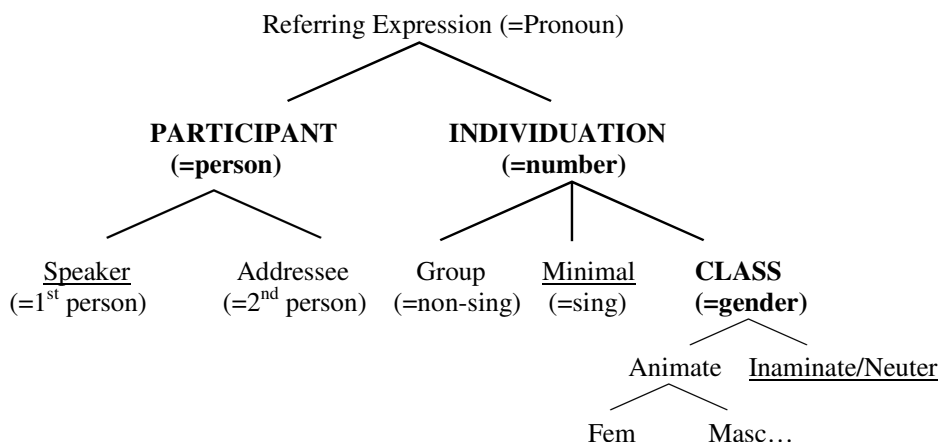
A change of referent in the discourse is marked via over material encoding the [Topic-shift] feature: (i) the overt pronoun *él* in (3) refers to the male character and *ella* to the female one; (ii) an NP is used when several antecedents are present, thus avoiding potential ambiguity, (4).

- (3) La última película que he visto es la de "El Ilusionista" [...] **Los protagonistas** son dos jóvenes que se conocen y se enamoran. **Él** es de clase baja, mientras que **ella** es de familia noble [...] [SPH, Spanish native, CEDEL2 corpus]
- (4) Un día **el príncipe** y **su prometida** acuden a ver el espectáculo [...] **El príncipe** sospecha de que **su prometida** le es infiel y [**pro**] manda a uno de sus secuaces a perseguirla [...] Al final, **el ilusionista** y **la chica** preparan su huida [...] **El príncipe** termina suicidándose y, al final, **el ilusionista** y **la chica** consiguen estar juntos. [SPH, Spanish native, CEDEL2 corpus]

2.2 Pronominal feature geometry

Phi-features (person, number, gender) have been traditionally treated as an unordered bundle (see section 3), but research on L1 acquisition shows that [1] [sing] and [3] [sing] [neut] are acquired earlier than [3] [sing] [\pm masc] (Harley & Ritter 2002a, 2002b; Hanson 2000). This stems from the 'Feature Geometry Analysis' (FGA) (Harley & Ritter 2002a, 2002b). UG provides a constrained and hierarchically organized set of pronominal features (Figure 1). Pronouns contain three main sets of features: PARTICIPANT (person) and its dependents, *Speaker* and *Addressee* represent 1st and 2nd person respectively; INDIVIDUATION (number) and its dependents, *Group* and *Minimal* correspond to non-singular (plural and dual) and singular number respectively; CLASS encodes gender and other class information. The underlined nodes (i.e., 1st person), *Minimal* (i.e., singular) and *Inanimate* (i.e., neuter) represent the default interpretation of the node.

Figure 1: Feature Geometry Analysis (Harley & Ritter 2002)



PARTICIPANT encodes two features: *Speaker* (1st person) and *Addressee* (2nd person), while 3rd person is unmarked. The crucial distinction between 1st / 2nd person (i.e., speech-act participants, deictic use of the pronoun) and 3rd person (i.e., anaphoric use of the pronoun) is not new (Benveniste 1971, Bloomfield 1933, Jespersen 1924). According to Harley & Ritter (2002a: 488) ‘The geometry ... captures the intuition that so-called 3rd person is in fact not a true personal form ... When the Participant node is absent, the underspecified Referring Expression node receives a so-called 3rd person interpretation’.

3 PREVIOUS L2 STUDIES

English-speaking learners of L2 Spanish acquire the *formal* (syntactic) properties that license null subjects from early stages (Liceras 1989, Lozano 2002a), but at the syntax-discourse interface learners show some persistent deficits with the *discursive* features ([Topic-continuity] and [Topic-shift]) (Liceras & Díaz 1999, Lozano 2002b, 2003, Montrul & Rodríguez-Louro 2006, Pérez-Leroux & Glass 1997, 1999). Previous research reports that the most common deficit is *overproduction* (an overt pronoun is redundantly produced in a topic-continuity context) and *underproduction* (a null pronoun is produced in a topic-shift continuity context).

Sorace (2004:144) observes that ‘aspects of grammar at the syntax-discourse interface are more vulnerable ... than purely syntactic ones’ (p. 143) and that ‘interfaces, because they are more complex than narrow syntax, are inherently more difficult to acquire’.

Recall that, unlike previous L2 research, I will present evidence showing that deficits at the syntax-discourse interface are *selective*, since not all persons are equally vulnerable.

4 PREDICTIONS

Based on previous L2 Spanish research and on the FGA, hypothesis (5) was formulated.

- (5) H₁: Deficits at the syntax-discourse interface do not affect the whole pronominal paradigm, but are rather *selective*: (i) speech-act participants (1st and 2nd person) and neutrals (3rd person inanimate) are impervious to deficits, but (ii) 3rd person animate (masculine and feminine) is vulnerable.

5 METHOD

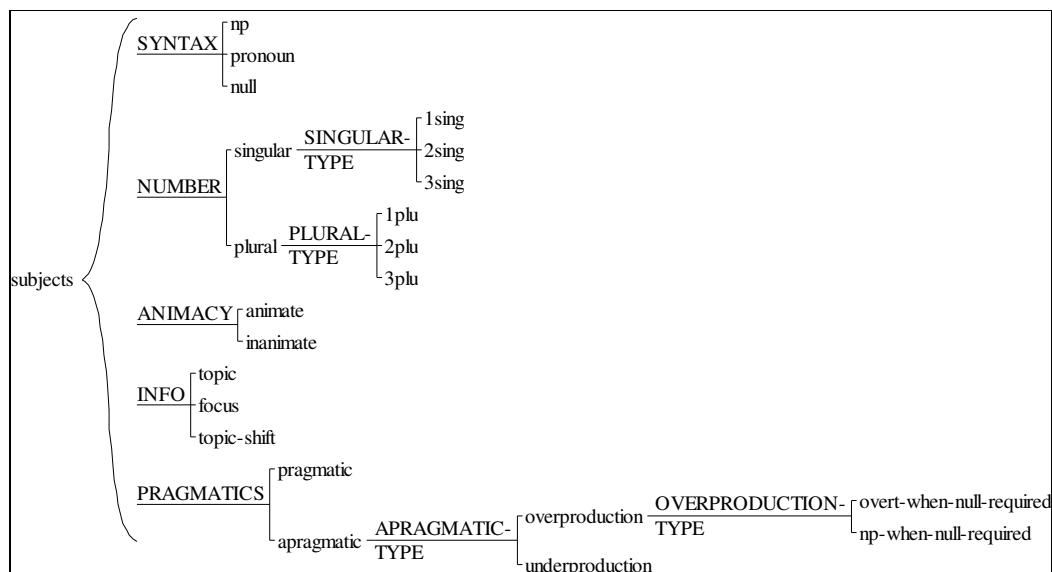
CEDEL2 (*Corpus Escrito Del Español L2*) is a written L1 English–L2 Spanish learner corpus (over 285,000 words) plus a native Spanish subcorpus (over 95,000 words) used for comparative purposes (more details at <http://www.uam.es/woslac/cedel2.htm>). Two groups of CEDEL2 learners were chosen (Table 1): upper-advanced and lower-advanced, as measured in the University of Winconsin College-Level Placement Test.

Group	N	Mean proficiency	Mean chronological age	Mean age first exposure	Mean years instruct	Mean stay in Spanish speaking country (months)
Upper-advanced	10	99%	35	14	8	29
Lower-advanced	10	93%	32	15	6	27
Spanish natives	12	n/a	37	n/a	n/a	n/a

Table 1: Learners' bio-data

The *UAM Corpus Tool* package (version 1.0; see <http://www.wagsoft.com/CorpusTool>) was used to annotate segments according to a previously defined scheme of linguistic features (tags), Figure 2, which was designed taking into account Harley & Ritter's (2002) FGA and previous research. Out of 32 texts (Table 2), each finite sentential subject was tagged.

Figure 2: Annotation scheme (tags)



Corpus	N texts	Total # words	Total # tags
Upper-advanced	10	8188	453
Lower-advanced	10	8521	528
Spanish natives	12	5954	299

Table 2: Analyzed texts

6 RESULTS

This study presents just a small set of results (further details can be found in Lozano, in press 2008).

6.1 Person and number of the subject

1st person singular (yo/pro 'I'): Production was pragmatically correct for all groups (over 98% for all groups), there being no significant differences between any of the groups ($\chi^2=2.234$, $df=2$, $p=0.327$), Table 3.

GROUP			PRAGMATICALITY 1 st sing		Total
			PRAGMATIC	UNPRAGMATIC	
LOW-ADV	Count		108	2	110
	% within GROUP		98.2%	1.8%	100.0%
UPP-ADV	Count		120	0	120
	% within GROUP		100.0%	0.0%	100.0%
SPANISH	Count		52	1	53
	% within GROUP		98.1%	1.9%	100.0%

Table 3: 1st singular

1st person plural (nosotros/nosotras/pro 'we'): Production of 1st plural was pragmatically correct (100% for all groups), Table 4.

GROUP			PRAGMATICALITY 1 st plural		Total
			PRAGMATIC	UNPRAGMATIC	
LOW-ADV	Count		42	0	42
	% within GROUP		100.0%	0.0%	100.0%
UPP-ADV	Count		27	0	27
	% within GROUP		100.0%	0.0%	100.0%
SPANISH	Count		22	0	22
	% within GROUP		100.0%	0.0%	100.0%

Table 4: 1st plural

2nd person singular (tú/pro 'you') and *2nd person plural (vosotros/vosotras/pro 'you all')*: All groups' production is pragmatically correct, though the frequencies are extremely low and no production by Spanish natives.

3rd person singular animate (NP/él/ella/pro 'NP/he/she'): While learners produce pragmatically incorrect forms (14.9% and 16.7%), the Spanish natives hardly produce any (Table 5). Learners significantly differs from Spanish natives (Fisher's Exact test: upper-advanced vs. natives: $p \leq 0.001$; lower-advanced vs. natives: $p \leq 0.001$).

			PRAGMATICALITY 3 rd sing animate		Total
			PRAGMATIC	UNPRAGMATIC	
GROUP	LOW-ADV	Count	175	35	210
		% within GROUP	83.3%	16.7%	100.0%
	UPP-ADV	Count	120	21	141
		% within GROUP	85.1%	14.9%	100.0%
	SPANISH	Count	97	1	98
		% within GROUP	99.0%	1.0%	100.0%

Table 5: 3rd singular animate

Overproduction with 3rd person singular animate, (6), implies redundant production of an overt pronoun (*él* ‘he’) to refer to the previous antecedent *mi novio de EEUU*. (The # symbol indicates pragmatic anomaly)

- (6) Cuando me faltaban dos semanas, mi novio de EEUU me visitó unos días para ver la ciudad que me encantaba tanto. Era la primera vez que #él salió de su país, por eso era un viaje importante. [CPB, upper-advanced, CEDEL2 corpus]

3rd person singular inanimate (*NP/ello/pro* ‘NP/it’): By contrast, learners show native-like behavior now, Table 6 (97.7% upper-advanced, 100% lower-advanced), similar to Spanish natives’ production (98.7%), with no significant differences (Fisher’s Exact test: upper-advanced vs. natives, $p=0.589$; lower-advanced vs. natives, $p=0.467$).

			PRAGMATICALITY of 3 rd sing inanim		Total
			PRAGMATIC	UNPRAGMATIC	
GROUP	LOW-ADV	Count	90	0	90
		% within GROUP	100.0%	0.0%	100.0%
	UPP-ADV	Count	43	1	44
		% within GROUP	97.7%	2.3%	100.0%
	SPANISH	Count	78	1	79
		% within GROUP	98.7%	1.3%	100.0%

Table 6: 3rd singular inanimate

Sentence (7) illustrates how the null pronoun (*pro*) is used pragmatically to refer to the 3rd person inanimate antecedent *El paseo de los ingleses*.

- (7) Cada día caminaba de mi apartamento a la universidad por “**El paseo de los ingleses**”. [*pro*] era un camino muy lindo con vistas de hoteles y también el mar azul y claro del Mediterráneo. [ARGL, upper-advanced, CEDEL2 corpus]

3rd person plural animate (*NP/ellos/ellas/pro* ‘NP/they’): Learners’ production of pragmatic forms (9.7% and 9.6%) is higher than natives’ production (5.6%), Table 7, yet non-significant (Fisher’s Exact Test: upper-advanced vs. natives: $p=0.348$; lower-advanced vs. natives: $p=0.394$).

			PRAGMATICALITY of 3 rd plu animate		Total
			PRAGMATIC	UNPRAGMATIC	
GROUP	LOW-ADV	Count	47	5	52
		% within GROUP	90.4%	9.6%	100.0%
	UPP-ADV	Count	102	11	113
		% within GROUP	90.3%	9.7%	100.0%
	SPANISH	Count	34	2	36
		% within GROUP	94.4%	5.6%	100.0%

Table 7: 3rd plural animate

3rd plural inanimate (NP/pro ‘they’): Learners show native-like behavior (Table 8) since they hardly produce any apragmatic tokens: 0% upper-advanced vs 0% natives, and 4.5% lower-advanced (which corresponds to only one residual token, which inflates the percentage) vs. 0% natives (non-significant, Fisher’s Exact Test $p=0.710$).

			PRAGMATICALITY of 3 rd plu inanim		Total
			PRAGMATIC	UNPRAGMATIC	
GROUP	LOW-ADV	Count	21	1	22
		% within GROUP	95.5%	4.5%	100.0%
	UPP-ADV	Count	9	0	9
		% within GROUP	100.0%	0.0%	100.0%
	SPANISH	Count	14	0	14
		% within GROUP	100.0%	0.0%	100.0%

Table 8: 3rd plural inanimate

6.2 Information status and pragmaticality of the subject

This section reports on the (a)pragmaticality of the information of status of the subject, i.e., whether the use of NP/overt/null is pragmatically correct.

Topic-continuity contexts: Both learner groups produce a significantly higher amount of apragmatic sentences (12.3% and 8.8%) than Spanish natives do (3%), Table 9: upper-advanced vs. natives: $\chi^2=5.621$, $df=2$, $p=0.018$; lower-advanced vs. natives: $\chi^2=11.269$, $df=2$, $p=0.001$. In (8) the learner is talking about *la madre*. A null pronoun is expected, as used in the first instance, to mark topic-continuity but the learner later uses two pragmatically redundant overt pronouns (*ella*).

- (8) **La madre** no puede hablar inglés pero [**pro**] es muy trabajadora. **#Ella** empieza a trabajar [...] **#Ella** no puede comunicar con esta familia. [SMM, Lower-Advanced, CEDEL2 corpus]

			PRAGMATICALITY		Total
			PRAG TOPIC	UNPRAG TOPIC	
GROUP	LOW-ADV	Count	299	42	341
		% within GROUP	87.7%	12.3%	100.0%
	UPP-ADV	Count	290	28	318
		% within GROUP	91.2%	8.8%	100.0%
	SPANISH	Count	159	5	164
		% within GROUP	97.0%	3.0%	100.0%

Table 9: Group * Pragmaticality of Topic

When a null pronoun is expected in topic-continuity contexts, errors can be of two types, overproduction of (i) an overt pronoun or (ii) an NP (Table 10, which shows the percentages out of the percentages of apragmatic topic in Table 9). All groups overproduce more overt pronouns than NPs, Table 10.

			OVERPRODUCTION TYPE		Total
			OVERT instead of NULL	NP instead of NULL	
GROUP	LOW-ADV	Count	33	9	42
		% within GROUP	78.6%	21.4%	100.0%
	UPP-ADV	Count	19	9	28
		% within GROUP	67.9%	32.1%	100.0%
	SPANISH	Count	3	2	5
		% within GROUP	60.0%	40.0%	100.0%

Table 10: Group * Overproduction type with Topic

Overproduction is shown in (9), where the use *ellos* to refer to its antecedent *los chicos* is pragmatically redundant. A null pronoun would be pragmatically adequate, as produced in the following clauses. Overproduction of NPs is also shown in (9), where the final instance of *los chicos* is redundant, since *los chicos* is already the topic in the discourse and, therefore, requires a null pronoun (*pro*) to mark topic-continuity.

- (9) Cuando me integré en el grupo, en realidad **los chicos** no podían cantar ni tocar muy bien. Sin embargo, poco a poco a lo largo del año, **#ellos** se mejoraron bastante y no sólo [*pro*] desarrollaron su grupo y sus talentos musicales, sino también [*pro*] crecieron como individuos. Tuvimos un retiro en que hablamos sobre las razones por las cuales **#los chicos** habían decidido participar en el grupo. [ELS, Upper-Advanced, CEDEL2 corpus]

Topic-shift contexts: Overt material (NP or overt pronoun) is required to mark a change in topic to prevent ambiguity between two or more potential antecedents. In underproduction errors, a null pronoun is produced when overt material is required. Learners' low percentages of underproduction (7.9% and 3.3%) correspond to just a few tokens, while Spanish natives never underproduce (0%), Table 11. Learner show some residual (but non-significant) underproduction (Fisher's Exact Test: upper-advanced vs. natives $p=0.57$, just non-significant; lower-advanced vs. natives $p=0.283$).

			PRAGMATICALITY		Total
			PRAG TOPICSHIFT	UNPRAG TOPICSHIFT	
GROUP	LOW-ADV	Count	88	3	91
		% within GROUP	96.7%	3.3%	100.0%
	UPP-ADV	Count	58	5	63
		% within GROUP	92.1%	7.9%	100.0%
	SPANISH	Count	47	0	47
		% within GROUP	100.0%	0.0%	100.0%

Table 11: Group * Pragmaticity of Topic-shift

7 DISCUSSION

Results support H₁ in (5) above:

- (i) *1st singular/plural* clearly show that learners show a pragmatic native-like production.
- (ii) *2nd singular/plural* show that learners use them in a native-like fashion. This result must be taken provisionally due to the low frequency observed in the CEDEL2 corpus.
- (iii) *3rd singular animate* clearly leads to deficits (significant differences) as well as *3rd plural animate* (non-significant). By contrast, learners do not show any deficits with *3rd singular/plural inanimate*.

H₁ is also supported topic-continuity contexts. Learners produce significantly more overt material (overt pronouns mostly and some NPs) to mark continuity in the discourse than Spanish natives. In topic-shift contexts, learners do produce some null pronouns when overt material is required, while Spanish natives do not, the difference being non significant.

Results on topic and topic-shift thus confirm the general hypothesis that advanced and end-state L2 learners show deficits at the syntax-discourse interface with pronominal subjects. In particular, English-speaking learners of L2 Spanish produce (i) a considerable proportion of overt pronouns in topic-continuity contexts and (i) a residual and non-significant amount of underproduction. Pragmatical errors are not across the board since they do not affect the whole pronominal paradigm, but are rather selective and affect only a subset of features in the paradigm (in particular, 3rd person animate), as predicted by the FGA, which is constrained by UG.

8 CONCLUSION

This study has used lower-advanced and upper-advanced data from the CEDEL2 corpus to show that deficits at the syntax-discourse interface are observable in the distribution of overt and null pronominal subjects in the discourse, as previous research as reported for advanced learners of L2 Spanish. But, unlike previous research, it has been shown that deficits are *selective*, affecting mainly 3rd person animate only, while the rest of the pronominal paradigm remains stable.

REFERENCES

- Alonso-Ovalle, L., Fernández-Solera, S., Frazier, L., and Clifton, C. 2002. Null vs. overt pronouns and the topic-focus articulation in Spanish. *Journal of Italian Linguistics* 14: 151-169.
- Benveniste, E. 1971. "Problems in General Linguistics". *The nature of pronouns*. Ed. E. Benveniste. Coral Gables, FL: University of Miami Press.
- Bloomfield, L. 1933. *Language*. New York: Holt, Rinehart and Winston.
- Hanson, R. 2000. "Pronoun acquisition and the morphological feature geometry". *Calgary Working Papers in Linguistics* 22: 1-14.
- Harley, H. and Ritter, E. 2002a. "Person and number in pronouns: A feature-geometric analysis". *Language* 78, 482-526.
- Harley, H. and Ritter, E. 2002b. "Structuring the bundle: A universal morphosyntactic feature geometry". *Pronouns: Grammar and Representation*. Eds. H.J. Simon, and H. Weise. Oxford: Elsevier.
- Jespersen, O. 1924. *The Philosophy of Grammar*. London: George Allen & Unwin.
- Liceras, J.M. 1989. "On some properties of the "pro-drop" parameter: looking for missing subjects in non-native Spanish". *Linguistic Perspectives on Second Language Acquisition*. Eds. S.M. Gass, and J. Schachter. Cambridge: Cambridge University Press. 109-133.
- Liceras, J.M., and Díaz, L. 1999. "Topic drop versus pro-drop: null subjects and pronominal subjects in the Spanish L2 of Chinese, English, French, German and Japanese speakers". *Second Language Research* 15: 1-40.
- Lozano, C. 2002a. "Knowledge of expletive and pronominal subjects by learners of Spanish". *ITL Review of Applied Linguistics* 135/6: 37-60.
- Lozano, C. 2002b. "The interpretation of overt and null pronouns in non-native Spanish". *Durham Working Papers in Linguistics* 8: 53-66.
- Lozano, C. in press 2008. "Selective deficits at the syntax-discourse interface: Evidence from the CEDEL2 corpus". *Representational Deficits in Second Language Acquisition*. Eds. Y.-I. Leung, N. Snape and M. Sharwood-Smith. Amsterdam: John Benjamins.
- Montrul, S. and Rodríguez-Louro, C. 2006. "Beyond the syntax of the Null Subject Parameter: A look at the discourse-pragmatic distribution of null and overt subjects by L2 learners of Spanish". *The Acquisition of Syntax in Romance Languages*. Eds. V. Torrens and L. Escobar. Amsterdam: John Benjamins. 401-418.
- Pérez-Leroux, A.T. and Glass, W.R. 1999. "Null anaphora in Spanish second language acquisition: probabilistic versus generative approaches". *Second Language Research* 15: 220-249.
- Pérez-Leroux, A.T. and Glass, W.R. 1997. "OPC effects on the L2 acquisition of Spanish". *Contemporary Perspectives on the Acquisition of Spanish, volume I: Developing Grammars*. Eds. A.T. Pérez-Leroux and W.R. Glass. Sommerville, MA: Cascadilla Press. 149-165.
- Sorace, A. 2004. "Native language attrition and developmental instability at the syntax-discourse interface: Data, interpretations and methods". *Bilingualism: Language and Cognition* 7: 143-145.
- Sorace, A. 2006. "Possible manifestations of shallow processing in advanced second language speakers". *Applied Psycholinguistics* 27: 88-91.
- University of Wisconsin, 1998. *The University of Wisconsin College-Level Placement Test: Spanish [(Grammar) Form 96M]*. Madison, WI: University of Wisconsin Press.