

DIRECT OBJECT CLITIC DOUBLING IN OT-LFG: A NEW LOOK AT RIOPLATENSE SPANISH

Bruno Estigarribia

Stanford University

Proceedings of the LFG05 conference

University of Bergen

Miriam Butt and Tracy Holloway King (Editors)

2005

CSLI Publications

<http://csli-publications.stanford.edu/>

Abstract

Spanish expresses the direct object argument of transitive clauses as a Direct Object Clitic, as a lexical or independent pronominal NP, or both (*Direct Object Clitic Doubling*). The latter structure presents an obvious puzzle to theories that assume one form or another of functional uniqueness. Although much research has been devoted to the structural representation and semantics of DOCLD, a rather natural question has been left uninvestigated in the linguistic literature: what drives these different types of expression?

In this paper I analyze the Rioplatense dialect (henceforth RSp), which is generally described as allowing CLD more freely than other dialects (including Standard Peninsular Spanish). Using the apparatus of OT, I investigate the relation between discourse structure, cross-linguistic markedness hierarchies, and formal expression of direct objects. My examples come from a corpus of four texts that I reference in the appendix, augmented by examples from the Internet which are in the public domain. Where no source is cited, the example is a constructed one, usually a modified example from the corpus.¹

1 Introduction

Direct Object clauses in Spanish can be classified in three groups². There are structures in which the direct object argument of a transitive verb is expressed by a lexical NP:

- | | |
|--|---|
| (1) (a) <Beto-42>
<i>Aldo pidió un mate</i>
Aldo asked a mate
Aldo ordered a “mate”. | (1) (b) <PerInv-14>
<i>De inmediato miró fijamente a sus compañeros</i>
Immediately looked.at-he fixedly A his-PL partners
He immediately looked fixedly at his partners. |
|--|---|

In other cases, the direct object argument is instantiated by an allotactic (Haiman 1985) direct object clitic (a *special clitic*, Zwicky 1977):³

- | | |
|--|--|
| (2) (a) <Beto-42>
<i>Roque lo miró fijo</i>
Roque CL he-looked-at fixed
Roque looked at him fixedly. | (2) (b) <Lig-45>
<i>Las debés tener</i>
CL you-must have
You sure have them. |
|--|--|

Finally, some structures have a direct object clitic that is coreferential with a lexical NP in the same clause:

- | | |
|--|---|
| (3) (a) <Lig-110>
<i>Yo <u>las</u> tenía guardadas <u>las</u> cartas⁴</i>
I CL had stored the letters
I had the letters stored. | (3) (b) <Beto-50>
<i>¿<u>La</u> vas a llamar <u>a</u> Marta?</i>
CL you-go to call A Marta
Are you going to call Marta? |
|--|---|

This paper focuses on Rioplatense Spanish, the dialect of the area around the Río de la Plata, including cities like Buenos Aires and Rosario, in Argentina, and Montevideo, in Uruguay.⁵ This dialect is generally described as allowing CLD more freely than Standard Peninsular Spanish, since in the former only inanimate direct objects can be doubled. The data were extracted from a corpus of seven conversations published in Ligatto (1996). These involve 15 participants, from 10 to 60 years of age, and two

¹ Special thanks to Joan Bresnan, T. Florian Jaeger, Peter Sells, Nigel Vincent, and the audience and organizers of LFG05 at Universitetet i Bergen, Norway.

² I will not address clitic doubling of indirect objects here.

³ The direct object clitics will be glossed “CL”, without indication of person, number or gender. Refer to the appendix for a table of DOCLs. Indirect Object Clitics, when they appear, will be glossed “IOCL”.

⁴ In all the examples in this paper, underlining signals the clitic and its associated constituent, boldface signals the clitic; when necessary, small capitals will mark focal stress. The glosses will be as transparent and non-technical as possible; however, the technical gloss A will be used for the animacy marker *a* which is obligatory before animate direct objects and is homonymous with the IO marker *a* and the preposition *a*.

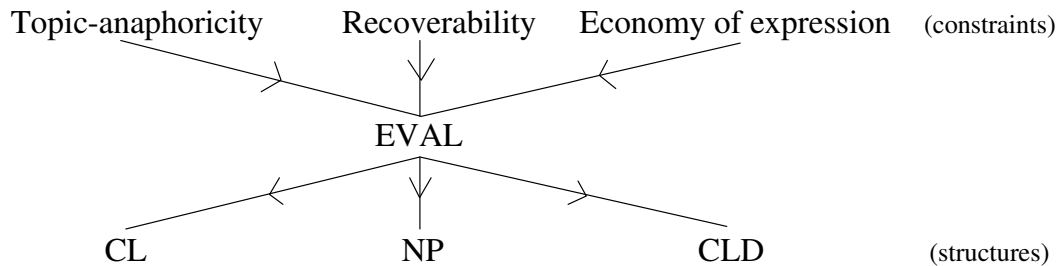
⁵ This dialect is also called Porteño Spanish or River Plate Spanish in the literature. It is spoken in a wider area than Buenos Aires, hence my choice of “Rioplatense”.

interviewers, all middle-class, born and raised in Buenos Aires. With the obvious exception of the ten-year old, all completed high-school at least. One participant was born in North-Eastern Argentina, but had lived in Buenos Aires since her adolescence. Data from a participant born in Spain were excluded. From the Ligatto corpus are also two short excerpts from a 1980's political talk-show, *Tiempo Nuevo*, involving two journalists and three politicians.

I supplemented this corpus with two short stories by Argentinian writer Roberto Fontanarrosa (1995), "Periodismo Investigativo" and "Beto", which contain fictional dialogues in the vernacular of Rosario. The characters in these stories are middle-class men and women in their forties and fifties, acquaintances and work colleagues in the first, old friends in the second. The origin of the examples is given between angled brackets, followed by the page number in the original text. The few constructed examples I used are marked as such (all ungrammatical examples are constructed).

Much research has tried to establish a structural representation for Direct Object Clitic Doubling structures that does not violate theoretical tenets. In LFG, for instance, Andrews (1990) proposed two lexical entries: one for stand-alone clitics with an obligatory PRED, and one for doubling clitics with an optional PRED and an ANIM + specification (which restricts doubling to animates and thus works for Peninsular Spanish but not for RSp).⁶ Also, researchers have tried to find factors that explain DOCLD's restricted distribution vis-à-vis the quasi-obligatory clitic doubling of indirect objects (see Estigarribia, forthcoming). However, these analyses are ultimately unsatisfactory because they do not address the question of why CLD exists and what functions it serves, especially considering the assumption, implicit in the name "clitic doubling", that the clitic is a pleonastic element.⁷ My working assumption here is that, in each structure, no lexical node is superfluous. Therefore, in CLD-clauses, both the presence of the DOCL and of an independent NP instantiating the same argument have to be justified. Note that this is consistent with LFG's principle of economy of expression. Bresnan (2001:147) suggests that the clitic's presence in the c-structure as a terminal node "still bears some kind of information not available elsewhere."⁸

My proposal is that three factors "conspire" to give the range of possible RSp transitive structures: Topic-anaphoricity (Bresnan 2001), associated with pronominal expression; Recoverability, associated with lexical expression; and Economy of expression, which prefers structures with as few lexical nodes as possible. This hypothesis is clearly in the spirit of Optimality Theory: a CLD construction would be the optimal result of conflicting pressures to give an argument a certain type of expression. CLD obtains when expression of both a clitic and a NP is the optimal response to independent constraints on expression of either argument.



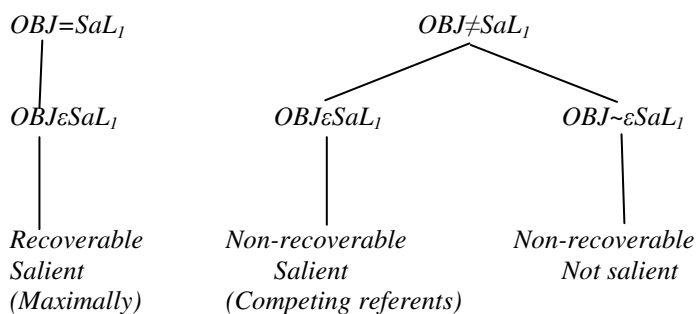
This analysis thus predicts under which conditions speakers will use a CLD structure. Furthermore, I will show that this approach can also explain two well-known phenomena: the so-called "obligatory" doubling of personal pronouns and the effect of animacy on DOCLD. Finally, the factorial typology predicts the existence of six types of languages. I will begin with my view of the input to the OT GEN module.

⁶ In derivational frameworks, the clitic and the lexical NP are assumed to originate in a "big DP" with a single theta-role, and Case is assigned to the "split" parts of the DP through different mechanisms (Belletti 2005).

⁷ Belletti (2005:31) speculates on a possible answer to this question, tentatively suggesting that "the clitic ultimately [contributes] to Case licensing of the noun phrase in Topic position." We'll see that this suggestion cannot work for Rioplatense, since topicality is not a necessary condition for CLD (section 6).

⁸ Bresnan also suggests that the clitic may be voided of nonredundant information, and in that case it would not contribute a separate node to the c-structure. This may indeed be the case for IO clitics, and would explain why clitic doubling of indirect objects is not restricted to referential arguments and is quasi-obligatory. In either case, economy-of-expression (in the classical LFG sense) is satisfied (see section 2 below).

Figure 1: Three possible relations of OBJ to SaL



I will turn now to the relation between topic-anaphoricity, recoverability, economy-of-expression and argument expression, and how to model them in OT to derive the distribution of DOCLD. In the tableaux that follow, Clitic-only structures will be represented as CL, NP-only structures as NP, and Clitic-doubling structures as CLD to render constraint evaluation easier, but of course c-structure representations are what are being evaluated.¹²

3 The constraints

3.1 Expression of DOCLs and topic-anaphoricity

According to Bresnan (2001), reduced pronominals are crosslinguistically specialized for topic-anaphoric uses. That is, when a reduced form is available, it is the form that will express topics. That DO clitics are topic-anaphoric in Spanish is shown by the following examples (these examples closely parallel those offered by Bresnan and Mchombo (1987) to demonstrate topic-anaphoricity of reduced forms in Chicheŵa):

Discourse topics:

(4) (a)

La hiena se comió al león. Habíendose_{lo} comido, se fue a San Francisco.
 The hyena ate the lion. Having-REFL-CL eaten went to San Francisco
The hyena ate the lion. Having eaten it (the lion), he went to S.F.

(b)

*La hiena se comió al león. *Habíendose comido a él, se fue a San Francisco.*
 The hyena ate the lion. Having-REFL eaten A him went to San Francisco
The hyena ate the lion. Having eaten him, he went to S.F.

One could claim that the ungrammaticality of the example above is due to the fact that personal pronouns can never appear without a clitic in object position¹³, but inanimate pronouns (which can appear without a doubling clitic) show the same pattern as animate pronouns:

(5) (a)

La hiena se comió el arroz. Habíendose_{lo} comido, se fue a San Francisco.
 The hyena ate the rice. Having-REFL-CL eaten went to San Francisco
The hyena ate the rice. Having eaten it (the rice), he went to S.F.

(b)

La hiena se comió el arroz. #Habíendose comido eso, se fue a San Francisco.
 The hyena ate the rice. Having-REFL eaten it/that went to San Francisco
The hyena ate the rice. Having eaten it/that (something other than the rice), he went to S.F.

¹² The reader can find these representations in many grammars of Spanish.

¹³ This situation is commonly known as “obligatory” CLD of personal pronouns. See below section 5.

Dislocated topics:

(6) (a)

Este arroz, la hiena se lo comió.

This rice, the hyena REFL CL ate

This rice, the hyena ate it.

(b)

**Este arroz, la hiena se comió eso.*

This rice, the hyena REFL ate it/that

(Intended) This rice, the hyena ate it.

Resumptive relativization:

(7) (a) <Lig-116>

Generalmente viste casas viejas que las arreglan...

Generally you-saw houses old that CL they-repair

In general, you saw old houses that people repair...

(b)

**Generalmente viste casas viejas que esas/a ellas arreglan.*

Generally you-saw houses old that those/A them they-repair

(Intended) In general, you saw old houses that people repair...

I take the preceding examples to be evidence in favor of the following constraint¹⁴:

OBJ_εSaL₁(cl): Assign one violation if DO in SaL₁ (salient) but not expressed by a clitic.¹⁵

That is, salient referents prefer clitic expression. Arguably, this constraint belongs to a family of constraints that harmonically align salient referents with the hierarchy of pronominals:

OBJ_εSaL₁(∅) >> OBJ_εSaL₁(aff) >> OBJ_εSaL₁(cl) >> OBJ_εSaL₁(weak) >> OBJ_εSaL₁(free)

Since RSp does not have DO verbal inflection, null anaphora¹⁶ or weak pronouns, all candidates will violate those constraints, and we can leave them out of the picture for modeling DOCLD.

3.2 Expression of DO NPs and recoverability

Lexical NPs differ from pronominals in having semantic content that allows them to:

- Introduce new referents and also new predications about referents
- “Point” to already established referents that satisfy the NPs’ lexical description

Pronominals can also “point” to referents that are accessible enough if they match the pronominal’s features. But if the referent is not immediately recoverable in the context of utterance (possibly through competition with other referents), then a lexical description (hence, a lexical NP) is needed. Having an operational definition of *recoverability* is crucial here (see Barbosa et al. 1998, Kuhn 2003, and Pesetsky 1998 for some problems in defining recoverability in an OT framework, and Buchwald et al. 2002 for a bidirectional approach): a DO is *recoverable* if OBJ=SaL₁, that is, it is the only element of SaL₁. If this is not the case, a form with lexical content will be needed:

OBJ≠SaL₁(NP): Assign one violation for every OBJ not identical with SaL₁ and not expressed by an NP

¹⁴ In this paper, constraints are verbally stated in the format recommended by McCarthy (2002, p.40).

¹⁵ Again, bear in mind that I am restricting myself to DOs here, but such a constraint could be applied to any GF, indeed to any argument of the predicate.

¹⁶ But see Masullo (2003) for a claim that null objects do exist in Rioplatense (they are very common in Andean Spanish and Basque Spanish).

Pesetsky (1998) and Kuhn (2003) assume that a recoverability constraint should be inviolable. Bresnan (2001) states that because of learnability considerations the input must be fully recoverable from the output. Clearly, recoverability constraints have a special status in OT. It is not even clear whether it should be a (rerankable) constraint at all, unless a language could show extreme unfaithfulness *systematically*. I believe that the problem lies there: allowing reranking in the factorial typology will yield *systematically* unrecoverable languages. But we want to allow reranking to minimize stipulations about the internal structure of the universal space of constraints. Unfortunately, I cannot tackle this issue here. I will assume a form of recoverability that is undominated without discussing whether this is a violation of the Methodological Principle of OT that states that explanation should be done by constraint interaction. This assumption will obviously affect the factorial typology, but I think it is a rather reasonable one.¹⁷

3.3 Economy of expression

Economy-of-expression constraints have been proposed under several guises in the OT literature. The version I will adopt here is the following:

***STRUC:** Assign one violation for each c-structure node (Aissen 2003).

Let's see now how these three constraints are ranked in RSp and how they interact to derive the basic distribution of CL-, NP-, and CLD-structures.

4 Constraint ranking and evaluation

In Rioplatense, our three constraints are ranked as follows: **OBJ≠SaL₁(NP) >> OBJεSaL₁(cl)>> *STRUC**. Tableau 1 summarizes evaluation of the three principal types of DO: Recoverable hence salient (OBJ=SaL₁), salient but not recoverable (OBJεSaL₁ but OBJ≠SaL₁), and not salient hence not recoverable (OBJ~εSaL₁). The letters 'n' and 'k' stand for the number of nodes in each structure, the relevant fact being that a CLD-structure will always violate *STRUC more times than either CL- or NP-structures.

Tableau 1: General distribution of transitive clause types

INPUT	CANDIDATES	(Recoverability)	(Topic-Anaphoricity)	(Economy)
		OBJ≠SaL ₁ (NP)	OBJεSaL ₁ (cl)	*STRUC
OBJ=SaL ₁ (hence OBJεSaL ₁)	CL ☺			*n + 1
	NP		*!	*n + k
	CLD			*n + k + 1!
OBJ≠SaL ₁ OBJεSaL ₁ ¹	CL	*!		*n + 1
	NP		*!	*n + k
	CLD ☺			*n + k + 1
OBJ~εSaL ₁ (hence OBJ≠SaL ₁)	CL	*!		*n + 1
	NP ☺			*n + k
	CLD			*n + k + 1!

We see that the basic distribution of direct transitive structures is as follows:

- CL-structures are optimal if the DO is salient and recoverable (maximally salient)
- NP-structures are optimal if the DO is not salient
- CLD-structures are optimal if the DO is salient but not recoverable (competition with other referents)

These generalizations are borne out by naturally occurring data.

¹⁷ A different approach, involving bidirectional optimization, is possible, but I will not explore it here. See Buchwald et al. 2002 and Kuhn 2003.

a) Maximally salient DOs

A clitic suffices (CL-structure) if the DO is the most salient referent

(8) <Beto-46>

Pero, uno se va con el problema. No lo dejás acá.

but one leaves with the problem no CL3SgM you.leave here

But you take the problem with you [when you go on vacation]. You don't leave it here.

In the above example, *el problema* 'the problem' is the only referential argument in the previous sentence (*uno* being a generic 'you'), which introduces it as (maximally) salient in the context.¹⁸

b) Non-salient referents

A NP suffices (NP-structure) if the DO is not a salient referent

(9) <Beto-42>

El Negro Moreira... dejó un par de cortados. Aldo le pidió un mate.

"Black" Moreira left a couple of espressos.with.milk. Aldo IOCL ordered a mate

"Black" Moreira left a couple of espressos [on the table]. Aldo ordered a mate [from him].

c) Salient but non-recoverable referents

CLD is the optimal candidate when the DO is salient but other equally salient referents exist.

(10) <Lig-115>

es mucho más fácil montar un jardín que-que un hospital o un consultorio-

is much more easy mount a garden than a hospital or a consultancy

It is much easier to start a kindergarten than a hospital or a private practice –

el jardín lo podés hacer poner con-con nada casi

the garden CL you-can make put with nothing almost

you can set up a kindergarten with almost no investment.

In this case, three salient referents are in competition, *un jardín* 'a kindergarten', *un hospital* 'a hospital', and *un consultorio* 'a private practice'. The Clitic Left-Dislocated structure (CLLD: Cinque 1900, Escobar 1997), a case of CLD, is used to pick out the intended referent.

We can see why it is crucial to clearly differentiate recoverability and salience as two independent notions in this model. Every salient argument will require a reduced form of expression, in this case a DO clitic, while every non-recoverable argument will require lexical content, that is, a NP or at least a free pronoun.¹⁹

Although the basic pattern is captured here, the actual distribution of DOCLD in discourse is more complicated. Morphological and semantic properties of the DO have an impact on whether a DOCLD-structure is optimal or dispreferred. I will now show that an independently grounded extension of this model successfully captures two well-known phenomena: "obligatory" doubling of personal pronouns and a seeming animacy restriction on DOCLD.

5 Personal pronouns and markedness

The literature on CLD considers it obligatory with personal pronouns because of the following contrasts:

¹⁸ Also the general topic of the exchange is how to avoid a certain problem one of the participants has, which contributes to the salient status of the referent 'the problem'. I am assuming here that some mechanism for calculation of salience is available, for instance, Centering Theory (Grosz, Joshi, and Weinstein 1995).

¹⁹ Remember that free pronouns are "both distributionally and prosodically indistinguishable from a full NP/DP" (Vincent 2001). The distribution of DOCLD with free pronouns will be derived by other constraints in the next section.

(11) (a) <PerInv-16>

*Sí, dejáme a mí que yo los conozco/ *Sí, dejá a mí que yo los conozco*
Yes leave-CL A me that I CL know Yes leave A me that I CL know
Leave it to me because I know them.

(12) (b) <Lig-124>

*Yo *(te) pregunto a vos*
I *(CL) ask A you
I am asking you.

(13) (c) <Lig-82>

*Yo *(la) veo a ella vender es bárbara yo admirada*
I *CL see A her sell is great I admiring
I see her sell... she's great, I'm all admiration.

(14) (d) <Lig-148>

**(me) llamó a mí Cámpora*
CL1Sg he.called A me Cámpora
Cámpora called [me]_{FOC}.

However, CLD-structures are not obligatory, since CL-structures are allowed:

(15) <Beto-50>

Ya te estaba extrañando
Already CL2Sg was missing
I missed you already.

Moreover, inanimate pronouns can appear without an accompanying clitic²⁰:

(16) (a) <Lig-81>

no soy yo en ese momento cuando estoy haciendo eso
no am I I that moment when I-am doing it/that
I'm not myself at the time when I am doing that.

(16) (b) <Lig-100>

el producto de la ignorancia es lo que favorece eso
the product of the ignorance is what favors it/that
It is the product of ignorance that favors that.

What needs to be accounted for is precisely the prohibition of NP-structures, when the NP is a free personal pronoun (that is, free and animate), not obligatoriness of CLD. This has been modeled in classical LFG as a morphological blocking effect (Andrews 1990). Now, the fact that only personal pronouns (which I interpret as [+anim, +pro]) present a morphological blocking effect indicates that there is an interaction with animacy that needs to be captured. What is the difference RSp makes in the treatment of animate and inanimate DOs?

In all varieties of Spanish the case of inanimate objects is indicated by the presence or absence of the preposition/marker *a*:

(17) (constructed example)

Le sacaste un botón a tu camisa
You took.out a button DAT your shirt.
You took a button off your shirt.

²⁰ These inanimate pronouns are homophonous with demonstratives. They could be argued to be demonstratives but they fill the inanimate gap in the paradigm of pronouns, and therefore, I take them to be inanimate pronouns, in accordance with prescriptive grammars of Spanish.

Note that *un botón* and *tu camisa* are both inanimate, but while the former is a DO and has no marking (except post-verbal position), the latter is an IO and is marked with *a*. However, Spanish requires all **referential animate** DOs to be marked with *a* too, and hence, case is not differentially marked for animates²¹ (example repeated from above):

(1) (b) < PerInv-14 >
De inmediato miró fijamente a sus compañeros
 Immediately looked.at-he fixedly A his-PL partners
He immediately looked fixedly at his partners.

Since the marker *a* is indistinguishable from the IO marker, in (1b) case is not unambiguously marked (locally). Avoidance of ambiguity was argued, for instance, by Donohue (1999) for Fore and is usually assumed in the processing literature to play a role in language production.²² Our constraint to avoid local ambiguity is:

***?CASE:** Assign one violation for every argument ambiguously marked for case.

Importantly, this constraint has an effect only in the case of animate DOs (remember that inanimate DOs are unambiguously marked by zero marking). But what is crucial here is that DO clitics, being distinct from IO clitics, can disambiguate marking in this case (García 1975). Therefore, structures with object clitics never violate ***?CASE**.

As Aissen (2003) notes, DOs are more marked the more definite they are, and therefore the following markedness scale (in the form of avoid constraints) applies:

***OBJ/PRO >> *OBJ/PN >> *OBJ/DEF >> *OBJ/SPEC >> *OBJ/NSPEC** (Aissen 2003, p. 445)

We can conjoin the ***?CASE** requirement with the definiteness hierarchy. The constraints that are the product of this conjunction are ranked above each separate constraint (local conjunction in the DO domain; see Smolensky 1995, McCarthy 2002 for discussion.):

***OBJ/PRO&*?CASE >> *OBJ/PN&*?CASE >> *OBJ/DEF NP&*?CASE >> *OBJ/SPEC NP&*?CASE >> *OBJ/NONSPEC NP&*?CASE**

Note the parallels with Differential Object Marking (Aissen 2003). To recapitulate, unambiguous case marking is satisfied by either an inanimate DO or an animate DO that is copresent with a coreferential clitic. For ease of presentation, let's collapse the non-pronominals into the category NP. The ranking for RSp is:

OBJ≠SaL₁(NP) >> OBJ=SaL₁(cl), *OBJ/PRO&*?CASE >> *STRUC, *OBJ/NP&*?CASE

where constraints within the same stratum are separated by commas and allowed to rerank. Tableau 2 below shows that the distribution of CLD/CL-structures and prohibition of NP-structures with animate pronouns follows without direct stipulation of a constraint that requires doubling of free personal pronouns.

The cases where the DO is salient (OBJεSaL₁) behave exactly as with a non personal pronoun argument, yielding a CL- or a CLD-structure. The evaluation differs for non-salient DOs: whereas in the general case (see tableau 1 above) a non-salient DO would enforce a NP-structure, in the case of personal pronouns such a situation requires a CLD-structure to avoid the local ambiguity of having marking of the DO by *a*. That is why NP-structures are not found with personal pronouns in Spanish. The crucial factor is that the constraint towards unambiguous case manifestation is satisfied by two independent means: it is satisfied by DO clitics, but it is also satisfied by any inanimate argument (which do not take *a* when DOs and take *a* when IOs). Hence, if we have an inanimate pronoun that is not salient, we revert to the general case where NP-structures are optimal (examples repeated from above):

²¹ Animate DOs in some cases may not be marked (nonspecific ones, for instance).

²² However, see Wasow (2002) for a sceptical view.

(16) (a) <Lig-81>
no soy yo en ese momento cuando estoy haciendo eso
 no am I I that moment when I-am doing it/that
I'm not myself at the time when I am doing that.

(16) (b) <Lig-100>
el producto de la ignorancia es lo que favorece eso
 the product of the ignorance is what favors it/that
It is the product of ignorance that favors that.

To summarize, we haven't explicitly built a constraint that requires doubling personal pronouns, but we have derived its effects from a constraint on case marking and from cross-linguistic, independently grounded markedness constraints.

Tableau 2: Distribution of transitive clauses with free personal pronouns as DOs

INPUT (all +Anim, +Pro)	CAND	(Recoverability)	(TopAnaph)	(AmbPro)	(AmbNP)	(Economy)
		OBJ≠SaL ₁ (NP)	OBJεSaL ₁ (cl)	*O/P & *?C	*O/NP & *?C	*STRUC
OBJ=SaL ₁	CL ☹					*n + 1
	NP		*!	*		*n + k
	CLD					*n + k + 1 _!
OBJ≠SaL ₁ OBJεSaL ₁	CL	*!				*n + 1
	NP		*!	*		*n + k
	CLD ☺					*n + k + 1
OBJ≠SaL ₁ OBJ~εSaL ₁	CL	*!				*n + 1
	NP			*!		*n + k
	CLD ☺					*n + k + 1

6 Animacy effect

Peninsular Spanish disallows DOCLD with inanimate DOs. Although this is not the case for RSp (contra Andrews 1990, Jaeggli 1986, Roberge 1990; see example (3a) above), some researchers claim that animates are doubled more frequently, or that inanimate DOCLD is less general in this dialect (Barrenechea and Orecchia 1977, Gutiérrez-Rexach 2000, Suñer 1988)²³, therefore acknowledging the presence of an *animacy effect*. However, the precise nature of this effect has never been stated. The OT model I propose derives a precise generalization: **animates, but not inanimates, can be doubled in contexts where the DO is not salient** (see Estigarribia, forthcoming).

[CONTEXT: B asks A for a prepaid phone card. "Marta" is B's girlfriend, not the topic of the exchange.]

(18) <Beto-50>

- A: ¿La vas a llamar a Marta?
 CL3SgF you.go to call A Marta
Are you going to call [Marta]_{FOC}?

- B: No querido... La voy a llamar a estamina de la que
 No dear CL3SgF I.go to call A this girl of CL that

hablábamos anoche

we.talked last.night

No my dear, I'm gonna call [this girl we were talking about last night]_{FOC}.

²³ Colantoni (2002) is the only author I know of that claims that inanimate DOCLD is more frequent than animate DOCLD.

The predictive strength and empirical adequacy of the model is demonstrated by the fact that the constraints already proposed account for this effect without further stipulation:

Tableau 3: Animacy effect on CLD

INPUT	CANDIDATES	(Recoverability)	(TopAnaph)	(AmbPro)	(AmbNP)	(Economy)
		OBJ≠SaL ₁ (NP)	OBJεSaL ₁ (cl)	*O/P & ?C	*O/NP & ?C	*STRUC
OBJ=SaL ₁ +Anim	CL ☹					* ⁿ⁺¹
	NP		*!		*	* ^{n+k}
	CLD					* ^{n+k+1} !
OBJ≠SaL ₁ OBJεSaL ₁ +Anim	CL	*!				* ⁿ⁺¹
	NP		*!		*	* ^{n+k}
	CLD ☹					* ^{n+k+1} !
OBJ≠SaL₁ OBJ~εSaL₁ +Anim	CL	*!				* ⁿ⁺¹
	NP				*!	* ^{n+k}
	CLD ☹					* ^{n+k+1} !
OBJ=SaL ₁ -Anim	CL ☹					* ⁿ⁺¹
	NP		*!			* ^{n+k}
	CLD					* ^{n+k+1} !
OBJ≠SaL ₁ OBJεSaL ₁ -Anim	CL	*!				* ⁿ⁺¹
	NP		*!			* ^{n+k}
	CLD ☹					* ^{n+k+1} !
OBJ≠SaL₁ OBJ~εSaL₁ -Anim	CL	*!				* ⁿ⁺¹
	NP ☹					* ^{n+k}
	CLD					* ^{n+k+1} !

The boldfaced cases are those in which the DO is not salient, and we see that animates and inanimates behave differently, the former preferring CLD-structures and the latter, NP-structures. Moreover, DOCLD with animates is optional with non-salient referents (as shown by the example below), and this is captured in the model by reranking constraints within the lowest stratum (see tableau 4):

(19) <Beto-48>
*¿Te enganachaste a la Sonia en lo del Pitu?*²⁴
 REFL you.hooked A the Sonia at.Pitu's
You picked up Sonia at Pitu's?

Tableau 4: Optionality of CLD with animates

INPUT	CAND	(Recoverability)	(TopAnaph)	(AmbPro)	(AmbNP)	(Economy)
		OBJ≠SaL ₁ (NP)	OBJεSaL ₁ (cl)	*O/P & ?C	*O/NP & ?C	*STRUC
OBJ≠SaL ₁ OBJ~εSaL ₁ +Anim	CL	*!				* ⁿ⁺¹
	NP				*!	* ^{n+k}
	CLD ☹					* ^{n+k+1} !

²⁴ This is an all-focus sentence.

		(Recoverability)	(TopAnaph)	(AmbPro)	(Economy)	(AmbNP)
INPUT	CAND	OBJ≠SaL ₁ (NP)	OBJεSaL ₁ (cl)	*O/P & ?C	*STRUC	*O/NP & ?C
OBJ≠SaL ₁	CL	*!			* ⁿ⁺¹	
OBJ~εSaL ₁	NP ⊙				* ^{n+k}	*!
+Anim	CLD				* ^{n+k+1} !	

7 Final ranking and synthesis of results

I checked the rankings obtained with the Gradual Learning Algorithm built into OTSoft (Hayes, Tesar and Zuraw 2003).²⁵ I assumed that the recoverability constraint was undominated (see discussion in 3.2) and I also assumed the cross-linguistic dominance relations derived from the conjunction of the overt case requirement with the object markedness hierarchy. I modeled free variation for the animate non-recoverable inputs (OBJ≠SaL₁, OBJ~εSaL₁, +Anim) by assigning winning frequencies of 0.5 to both the NP- and CLD-structure candidates. The stochastic grammar found was:

OBJ≠SaL₁(NP) = 142
 *OBJ/PRO & ?CASE = 122
 OBJεSaL₁(cl) = 108
 OBJ/NP&?CASE = 96.27
 *STRUC = 95.73

The grammar correctly predicted all winning candidates and modeled free variation in the animate non-recoverable case as output frequencies of 0.42 for the NP-structure and 0.581 for the CLD-structure, very close to the theoretical 0.5 frequencies. The average error per candidate was less than 0.3%. In what follows, I will use the ranking in stratal form, and abbreviated constraint names:

OBJ≠SaL₁(NP) >> *OBJ/PRO & ?CASE >> OBJεSaL₁(cl) >> *OBJ/NP&*?CASE, *STRUC
 that is
 Rec >> AmbPro >> TopAnaph >> AmbNP, ECON

I will also use a binary feature vector representation for the inputs for readability: [±sal, ±rec, ±anim, ±pro].

7.1 Optimization of the different cases

A) Salient recoverable DOs (maximally salient DOs = OBJ=SaL₁) : Clitic structure is optimal.

[+sal, +rec, +anim, +pro]	REC	AmbPro	TopAnaph	AmbNP	ECON
CL ⊙					* ⁿ⁺¹
NP		*!	*		* ^{n+k}
CLD					* ^{n+k+1} !

[+sal, +rec, +anim, -pro]	REC	AmbPro	TopAnaph	AmbNP	ECON
CL ⊙					* ⁿ⁺¹
NP			*!	*	* ^{n+k}
CLD					* ^{n+k+1} !

²⁵ Initial rankings: all constraints = 100; 50000 learning trials; initial plasticity = 2; final plasticity = 0.002; a priori rankings differ by 20; grammar tested 2000 times.

[+sal, +rec, -anim, +pro]	REC	AmbPro	TopAnaph	AmbNP	ECON
CL ☹					*n + 1
NP			*!		*n + k
CLD					*n + k + 1!

[+sal, +rec, -anim, -pro]	REC	AmbPro	TopAnaph	AmbNP	ECON
CL ☹					*n + 1
NP			*!		*n + k
CLD					*n + k + 1!

This result is optimal for every possible ranking of the constraints: the construction is optimal in every possible language.

B) Salient non-recoverable DOs (OBJ≠SaL₁, OBJεSaL₁): CLD-structure is optimal.

[+sal, -rec, +anim, +pro]	REC	AmbPro	TopAnaph	AmbNP	ECON
CL	*!				*n + 1
NP		*!	*		*n + k
CLD ☹					*n + k + 1

[+sal, -rec, +anim, -pro]	REC	AmbPro	TopAnaph	AmbNP	ECON
CL	*!				*n + 1
NP			*!	*	*n + k
CLD ☹					*n + k + 1

[+sal, -rec, -anim, +pro]	REC	AmbPro	TopAnaph	AmbNP	ECON
CL	*!				*n + 1
NP			*!		*n + k
CLD ☹					*n + k + 1

[+sal, -rec, -anim, -pro]	REC	AmbPro	TopAnaph	AmbNP	ECON
CL	*!				*n + 1
NP			*!		*n + k
CLD ☹					*n + k + 1

Reranking in this case may lead to other optimal candidates in other languages (see section 8 on Factorial Typology).

C) Non-salient, non-recoverable DOs (OBJ≠SaL₁ OBJ~εSaL₁):

a) CLD is optimal with personal pronouns.

[-sal, -rec, +anim, +pro]	REC	AmbPro	TopAnaph	AmbNP	ECON
CL	*!				*n + 1
NP		*!			*n + k
CLD ☹					*n + k + 1

Reranking may lead to a different outcome in different languages.

b) CLD is optional with non-pronominal animates (NP construction is also possible).

[-sal, -rec, +anim, -pro]	REC	AmbPro	TopAnaph	AmbNP	ECON
CL	*!				* ⁿ⁺¹
NP ☹				*	* ^{n+k}
CLD ☹					* ^{n+k+1}

Reranking may lead to a different outcome in different languages.

c) NP construction is optimal with inanimates.

[-sal, -rec, -anim, +pro]	REC	AmbPro	TopAnaph	AmbNP	ECON
CL	*!				* ⁿ⁺¹
NP ☹					* ^{n+k}
CLD					* ^{n+k+1} !

[-sal, -rec, -anim, -pro]	REC	AmbPro	TopAnaph	AmbNP	ECON
CL	*!				* ⁿ⁺¹
NP ☹					* ^{n+k}
CLD					* ^{n+k+1} !

This configuration obtains for every possible language.

As we see in the analysis above, the cases where either CL or NP structures (undoubled structures) are optimal cannot yield a different result under reranking. Note that it may seem that this is a strange situation, at first glance. But the constraint enforcing reduced expression of topics is formulated exclusively in terms of clitics. If we included the whole hierarchy of reduced pronominal forms, the winning CL candidates would violate higher ranked constraints requesting zero and affixal inflection, and lower ranked ones requesting weak and independent pronominals. Therefore, no candidate is completely unmarked and faithful (an undesirable situation from the theoretical point of view).

8 Factorial typology

8.1 Language Space and Variation Space

With 5 constraints, the number of logically possible grammars is 120. Using OTSoft, there were 6 different output patterns, represented in Table 1 below.

Therefore, the factorial typology predicts that:

- All recoverable DOs are expressed by CL structures in every possible language.
- All the inanimate non-salient DOs are expressed by NP structures in every possible language.
- The differing outputs correspond to salient non-recoverable DOs and to non-salient animate DOs.

Note that the output correctly predicts that the implicational hierarchies of definiteness and animacy (and possibly topicality, if operationalized through salience) will be respected cross-linguistically:

- If [-pro] allows CLD, then, *ceteris paribus* [+pro] allows CLD;
- If [-anim] allows CLD, then, *ceteris paribus* [+anim] allows CLD;
- If [-sal] allows CLD, then, *ceteris paribus* [+sal] allows CLD.

Table 1: Factorial typology for direct object clitic doubling

Language Type				I Rioplatense	II Rioplatense (optionally)	III Bulgarian	IV Peninsular Spanish	V Kichaga	VI French, Italian
Recoverable	Salient	+Anim	+Pro	CL					
			-Pro						
		-Anim	+Pro						
			-Pro						
Non-Recoverable	Salient	+Anim	+Pro	CLD					
			-Pro						
		-Anim	+Pro						
			-Pro						
	Non-salient	+Anim	+Pro	NP					
			-Pro						
		-Anim	+Pro						
			-Pro						

8.2 Possible rankings and possible languages

Language 1

REC >> TopAnaph, AmbPro >> ECON >> AmbNP

Language 2

REC >> TopAnaph, AmbPro >> AmbNP >> ECON

Rioplatense Spanish behaves optionally like Language 1 or Language 2, since CLD/NP are both possible for [-sal, -rec, +anim, -pro]. In that case, given that variation here is modeled as reranking within a stratum, we get the ranking given in section 7 above, and repeated here for convenience:

REC >> TopAnaph, AmbPro >> AmbNP, ECON

Note that, since the economy-of-expression constraint *STRUC is ranked very low, this language has a very widespread CLD pattern.

Language 3

REC >> TopAnaph >> ECON >> AmbPro >> AmbNP

Language 3 is a language that expresses non-salient DOs as NP-structures, salient recoverable DOs as CL, and salient non-recoverable ones as CLD. Bulgarian (Jaeger and Gerassimova 2002) is possibly such a language. Since *STRUC is ranked above the unambiguous marking markedness constraints, in such a language unambiguous case marking constraints are unnecessary, and salience/topicality is the major factor that drives CLD.

Language 4

REC >> AmbPro >> AmbNP >> ECON >> TopAnaph

Language 4 only has CLD with animates. This is what is usually claimed about Peninsular Spanish. In such a language salience/topicality would play no role (“emergence of the unmarked” effects aside) in clitic expression. Unambiguous marking of cross-linguistically marked DOs is the most relevant factor here, and so the pattern of CLD reflects animacy very accurately.

Language 5

REC >> AmbPro >> ECON >> TopAnaph, AmbNP

Language 5 has CLD only for non-recoverable personal pronouns (both salient and not). Kichaga as reported by Bresnan and Moshi (1990) could be such a language. AmbNP plays no role in this case.

Language 6

REC >> ECON >> TopAnaph, AmbPro >> AmbNP

Language 6 has no CLD. Economy-of-expression is ranked very high, above all the other markedness constraints, dominated only by recoverability. In this language, unambiguous marking of objects does not play a role either.

Since it is usually claimed that French and Italian do not have CLD, this could be the ranking underlying them. However, French and Italian do not fit this picture since they do have CLLD,²⁶ that is Clitic Left-Dislocated structures, which I analyze as a particular case of CLD (Belletti 2005, Estigarribia forthcoming). However, they do not have CLD of in-situ direct objects and this suggests that other constraints may be needed for the model to make more accurate cross-linguistic predictions.

This typology allows us to make interesting observations about these possible languages. In all of them, non-recoverability of information is a major divide between possible expressions of arguments (but this result was somewhat built into the analysis; see section 9, Conclusions). A language of type 6 only differentiates recoverable and non-recoverable arguments. The other five types make additional distinctions.

Only in one of them (Language 3), the presence of a clitic unambiguously encodes salience/topicality, although this is a prevalent hypothesis in the literature for Spanish CLD. In fact, only in two language types (3 and 6), the presence of a clitic encodes one only factor: recoverability in 6, salience/topicality in 3. In the remaining 4 types, the presence of a clitic will be determined by a more complex combination of animacy, pronominality, and discourse structure.

9 Conclusions

This paper showed the route towards a complete cross-linguistic OT analysis of Clitic Doubling, based on data from Rioplatense Spanish. The dimensions of animacy, definiteness and discourse structure were shown to interact to determine the possible patterns of CLD in the world's languages. The proposal advanced here was related to functional and typological results on pronominal expression and markedness results with respect to Differential Object Marking. This, I believe, gives this contribution a solid grounding in ongoing OT research.

Also, DO clitics have been shown not to be associated with one particular function, but rather as providing means for expressing independent requirements: as salience/topicality markers or as overt markers of case in surface-ambiguous predicate structures. The crucial point on which the analysis hinges is that recoverability and salience are factors that can be teased apart from one another. Even though *recoverability* as defined here entails *salience*, the converse is not true, and DOCLD emerges precisely when *salience* obtains but *recoverability* doesn't. This situation has not been fully exploited in previous analyses.

I chose not to deal directly with structural considerations, and therefore the role of structural descriptions (like LFG) was minimized. Far from being a weakness of the analysis, I consider this to be one of its strengths. The results exposed are maximally independent from theory-internal reasoning.

Several problems with current formalizations of OT syntax were spotted. The most salient is the issue of how to deal with ellipsis and recoverability in such an approach. Methodologically speaking, it would be desirable that such effects were a result of the interaction of violable constraints. However, so far such a solution has proven elusive, and some sort of inviolable recoverability principle still needs to be assumed. Bidirectional optimization may offer a solution to this quandary (Buchwald et al. 2002, Kuhn 2003), but this is a field that needs to be explored in more depth.

²⁶ My thanks to Nigel Vincent for this observation.

Insofar as the input is concerned, the addition of a Saliency List with a range of possible positions (degrees of saliency) may be useful to model also secondary topic phenomena, of the sort discussed by Dalrymple and Nikolaeva (2005).

A potential quirk in the analysis is the received knowledge that CLD is a cline towards grammatical agreement. If this is the case, then it should be possible to get CLD across the whole spectrum of possibilities. But, as we have seen, inanimate non-salient DOs are expressed by NP-only structures in every possible language. One way of explaining this and making the analysis more powerful is to include the markedness scales (definiteness and animacy) in full. That way, economy of expression would be allowed to interact in finer-grained ways with requirements for overt case marking.

APPENDIX

Direct Object Clitics in Rioplatense Spanish

	Masculine	Feminine
1 Sg	me	
2 Sg	te	
3 Sg	lo	la
1 Pl	nos	
2 Pl	los	las
3 Pl	los	las

Corpus references

In the examples, the texts are referred to by the abbreviations between angle brackets, followed by the page number in the original.

<PerInv> Fontanarrosa, Roberto. 1995. “Periodismo Investigativo”, in *La mesa de los galanes y otros cuentos*. Buenos Aires: Ediciones de la flor. 7-16.

<Beto> Fontanarrosa, Roberto. 1995. “Beto”, in *La mesa de los galanes y otros cuentos*. Buenos Aires: Ediciones de la flor. 37-50.

[*La mesa de los galanes* is a book of comic short stories reproducing very accurately the vernacular dialogues of the rioplatense area where RSp is spoken].

<Lig> Ligatto, Dolores. 1996. *Matériau pour l'étude de l'espagnol parlé : la variante argentine*. Limoges: Pulim. [Transcription and translation into French of several corpora of naturally occurring conversation and television interviews from the 1980s, all of them representing RSp]

<Quino> Quino (Joaquín Salvador Lavado). 1997. *Toda Mafalda*. Buenos Aires: Ediciones de la flor. [Comic strips of the 60s, 70s and 80s]

References

- Aissen, Judith. 2003. Differential Object Marking: Iconicity vs. Economy. *Natural Language and Linguistic Theory* 21, 435-83.
- Andrews, Avery. 1990. Unification and Morphological Blocking. *Natural Language and Linguistic Theory* 8: 507-557.
- Barbosa, Pilar, Danny Fox, Paul Hagstrom, Martha McGinnis and David Pesetsky (eds.). 1998. Introduction. In *Is the Best Good Enough? Optimality and Competition in Syntax*. Cambridge, MA – London: MIT.
- Barrenechea, Ana María and Teresa Orecchia. 1977. La duplicación de objetos directos e indirectos en el español hablado en Buenos Aires. *Estudios sobre el español hablado en las principales ciudades de América*, ed. by Lope Blanch, Juan M., 351-381. México: Universidad Autónoma de México.
- Belletti, Adriana. 2005. Extended doubling and the VP periphery. *Probus* 17. 1-35.
- Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell.

- Bresnan, Joan. 2001. The Emergence of the Unmarked Pronoun. In *Optimality-theoretic Syntax*, edited by Geraldine Legendre, Jane Grimshaw, and Sten Vikner, 113 - 142. Cambridge, MA: The MIT Press.
- Bresnan, Joan and Sam A. Mchombo. 1987. Topic, Pronoun and Agreement in Chicheŵa. *Language* 63. 741-782.
- Bresnan, Joan and Lioba Moshi. 1990. Object asymmetries in comparative Bantu syntax. *Linguistic Inquiry* 21. 147-185.
- Buchwald, Adam, Oren Schwartz, Amanda Seidl, and Paul Smolensky. 2002. Recoverability Optimality Theory: Discourse Anaphora and Bidirectional Optimization. In Johan Bos, Mary Ellen Foster and Colin Matheson (eds.): *Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002)*. 37-44.
- Cinque, Guglielmo. 1990. *Types of A'-Dependencies*. Cambridge: MIT Press.
- Colantoni, Laura. 2002. Clitic Doubling, Null Objects and Clitic Climbing in the Spanish of Corrientes, in Gutierrez-Rexach, J. (ed.). *From Words to Discourse: Trends in Spanish Semantics and Pragmatics*. Oxford: Elsevier.
- Dalrymple, Mary, and Irina Nikolaeva. 2005. Agreement and discourse function. Paper presented at LFG05, Universitetet i Bergen.
- Donohue, Cathryn. 1999. Optimizing Fore case and word order. Stanford University, ms.
- DuBois, John W. 1980. Beyond Definiteness: The Trace of Identity in Discourse. In Wallace L. Chafe (ed.), *The Pear Stories. Cognitive, Cultural and Linguistic Aspects of Narrative Production*. Series Advances in Discourse Processes, Vol III. Norwood, NJ: Ablex.
- Escobar, Linda. 1997. Clitic left dislocation and other relatives. In Elena Anagnostopoulou, Henk van Riemsdijk and Frans Zwarts, *Materials on Left Dislocation*. Amsterdam: John Benjamins. 233-273.
- Estigarribia, Bruno. Forthcoming. Why clitic doubling? A functional analysis for Rioplatense Spanish. *Selected proceedings of the 8th Hispanic Linguistics Symposium and 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages*. Somerville: Cascadilla Press.
- García, Erica. 1975. *The Role of Theory in Linguistic Analysis: The Spanish Pronoun System*. Amsterdam: North Holland.
- Grosz, Barbara, Aravind Joshi, and Scott Weinstein. 1995. Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics* 21. 203-225.
- Gutiérrez-Rexach, Javier. 2000. The Formal Semantics of Clitic Doubling. *Journal of Semantics* 16. 315-380
- Haiman, John. 1985. *Natural Syntax: Iconicity and Erosion*. Cambridge: Cambridge University Press.
- Hayes, Bruce, Bruce Tesar, and Kie Zuraw (2003) "OTSoft 2.1," software package, <http://www.linguistics.ucla.edu/people/hayes/otsoft/>.
- Jaeger, T. Florian and Veronica A. Gerassimova. 2002. Bulgarian Word Order and the Role of the Direct Object Clitic in LFG. In *Proceedings of the LFG02 Conference*, Stanford, CA: CSLI.
- Jaeggli, Osvaldo A. 1986. Three Issues in the Theory of Clitics: Case, Doubled NPs, and Extraction. In Hagit Borer (ed.), *Syntax and Semantics 19: The Syntax of Pronominal Clitics*. New York: Academic Press. 15-42.
- Kuhn, Jonas. 2003. *Optimality-Theoretic Syntax – A Declarative Approach*. Stanford, CA: CSLI.
- Masullo, Pascual. 2003. Cliticless definite object drop in River Plate Spanish, paper presented at LSRL XXXIII, Indiana University.
- McCarthy, John J. 2002. *A Thematic Guide to Optimality Theory*. Cambridge, UK: Cambridge University Press.
- Pesetsky, David. 1998. In Pilar Barbosa, Danny Fox, Paul Hagstrom, Martha McGinnis and David Pesetsky (eds.), *Is the Best Good Enough? Optimality and Competition in Syntax*. Cambridge, MA – London: MIT.
- Roberge, Yves. 1990. *The Syntactic Recoverability of Null Arguments*. Kingston – Montreal: McGill-Queen's University Press.
- Sells, Peter. The INPUT and Faithfulness in OT Syntax. 2003. In Jennifer Spenader, Anders Eriksson and Oesten Dahl (ed.) *Proceedings of the workshop on Variation within Optimality Theory*, University of Stockholm, 92-101.
- Smolensky, Paul. 1995. On the Internal Structure of the Constraint Component Con of UG. Manuscript available at the Rutgers Optimality Archive <http://roa.rutgers.edu/index.php3>.
- Suñer, Margarita. 1988. The role of agreement in clitic-doubled constructions. *Natural Language and Linguistic Theory* 6: 391-434.

- Vincent, Nigel. 2001. Competition and correspondence in syntactic change: null arguments in Latin and Romance. In *Diachronic Syntax*, ed. by Susan Pintzuk, George Tsoulas, and Anthony Warner. Oxford: Oxford University Press. 25-50.
- Wasow, Tom. 2002. *Postverbal Behavior*. Stanford: CSLI.
- Zwicky, Arnold 1977. *On clitics*. Bloomington: Indiana University Linguistics Club.