

The Morphology of Romance Clitic Clusters*

Diego Pescarini

University of Padova

In this paper I suggest that synthetic clitic clusters (i.e. sequences of clitics displaying a mismatch between their morphological form and their syntactic functions) are due to a constraint on vocabulary insertion (OCP) disallowing the same Vocabulary Item to be inserted more than once in the same cluster.

Firstly, I discuss the standard analysis, according to which violations of this parameter would trigger post-syntactic operations, allowing the insertion of unexpected exponents. Secondly, I offer a different hypothesis suggesting that a clitic blocked by the OCP is *automatically* replaced by an elsewhere exponent, i.e. the less specific clitic in the Vocabulary.

1 INTRODUCTION

This paper deals with Romance clitic clusters; in particular I analyse synthetic clusters, i.e. sequences of clitics displaying a mismatch between their morphological form and their syntactic functions (Bonet 1991, 1995; Harris 1994, 1997). For example, in Spanish (1) a cluster formed by a dative /le/ and an accusative /lo/ is not realized as /le lo/ but as /se lo/ where a reflexive clitic *se* appears instead of the dative one. Other examples of synthetic clusters are shown in (2).

- | | | | | | | | | |
|-----|-----|------------|---|------------|---|----------------------------|--------------------|-------------------|
| (1) | | <i>le</i> | + | <i>lo</i> | = | <u><i>se</i></u> <i>lo</i> | (<i>*le lo</i>) | (Spanish) |
| | | 3.dat | | 3.acc | | | | |
| (2) | (a) | <i>le</i> | + | <i>lo</i> | = | <u><i>le</i></u> <i>lo</i> | (<i>*le lo</i>) | (Italian) |
| | | 3.dat.f | | 3.acc | | | | |
| | (b) | <i>li</i> | + | <i>lu</i> | = | <u><i>nu</i></u> | (<i>*li lu</i>) | (Castrovillari) |
| | | 3.dat | | 3.acc | | | | |
| | (c) | <i>li</i> | + | <i>lu</i> | = | <u><i>bi</i></u> <i>lu</i> | (<i>*li lu</i>) | (Logudoro, Sard.) |
| | | 3.dat | | 3.acc | | | | |
| | (d) | <i>lez</i> | + | <i>li</i> | = | <u><i>zi</i></u> | (<i>*lez li</i>) | (Barceloní, Cat.) |
| | | 3.acc.f.pl | | 3.dat | | | | |
| | (e) | <i>si</i> | + | <i>si</i> | = | <u><i>ci</i></u> <i>si</i> | (<i>*si si</i>) | (Italian) |
| | | reflexive | | impersonal | | | | |

I will claim that these patterns are predictable and my goal is to develop a unitary account of them in the framework of Distributed Morphology (Halle & Marantz 1993, 1994). In sections 2 and 3 I will discuss two alternative hypotheses: an operation-based approach vs an item-based one.

The main tenet of my analysis is that these phenomena are due to a markedness constraint disallowing the same morpheme (namely the same vocabulary item) to be inserted more than once in the same cluster (Ackema 2001; Grimshaw 1997; Menn & MacWhinney 1984; Yip 1998; Neelman & Van der Koop 2004; Ortmann & Popescu 2000). I will refer to this principle as a

* For helpful comments and discussions on this material I thank Jonathan Bobaljik and Andrea Calabrese.

(9)		-f	+f
	accusative	<i>lo(s)</i>	<i>la(s)</i>
	dative	<i>le(s)</i>	

(10) [gender] → ∅ / [dat]

After post-syntactic operations like (10), vocabulary items are inserted into the slots of the template. The match between a slot of the template and a vocabulary item is ruled by the Subset Principle (Halle & Marantz 1993, 1994) which states that

(11) the features representing a vocabulary item are a subset of the features characterizing the slot; the most specific item (among the underspecified ones) will be inserted.

The derivation (12) describes the insertion of Vocabulary Items in the structure of the Spanish cluster in (1). The dative clitic (on the left) is impoverished, then Vocabulary Insertion takes place on the basis of the feature representations. But the output of this derivation does not correspond to the real form of the cluster (1). Indeed, in the dative clitic the /l/ exponent and the plural marker are replaced by a reflexive /s/.

(1) *les* + *lo* = *selo* (**leslo*) (Spanish)
 3.+pl.dat 3.-f.-pl.acc

(12)	template	pers	gen	num	case	+	pers	gen	num	case
	features	3	+/-f	+pl	dat		3	-f	-pl	acc
	impoverishment	3	∅	+pl	dat		3	-f	-pl	acc
	Vocabulary Insertion	<i>l</i>	<i>e</i>	<i>s</i>	-		<i>l</i>	<i>o</i>	-	-

In accordance with the discussion above, I suggest that the derivation (12) is blocked by OCP that does not allow the insertion of two *l* exponents in the same cluster.

Now I will discuss two hypotheses on the strategy adopted by the Morphological component to repair OCP violations.

2 OCP TRIGGERS REPAIRS

The first hypothesis is consistent with other Distributed Morphology analyses based on post-syntactic operations. Indeed, I suggest that, when OCP is violated – as in the case (1) – the morphological component modifies again the feature representation of the cluster allowing the insertion of a different exponent. These modifications are captured by language-specific operations of impoverishment that, following Calabrese (1994), I will call ‘repairs’.

For instance in Spanish (13) I suggest that the [number] feature of the dative clitic is impoverished - as shown in (14) - allowing the insertion of the reflexive marker *se* that is a bare 3rd person marker, a vocabulary item that does not need further gender and number specifications.

(13) template features impoverishment insertion repair insertion2	pers	gen	num	case	+	pers	gen	num	case
	3	+/-f	+pl	dat		3	-f	-pl	acc
	3	∅	+pl	dat		3	-f	-pl	acc
	*l OCP!	∅	+pl	dat		<i>l</i>	<i>o</i>	-	-
	3	∅	∅	-					
	<i>s</i>	<i>e</i>	-	-					

(14) [number] → ∅

Some American dialects bring evidence in favour of this kind of analysis since they replace the dative clitic with the reflexive exponent as standard Spanish, and, at the same time, they “move” the plural marker on the accusative clitic (Kany 1945). This phenomenon – described in (15) – is called parasitic or floating plural. According to my hypothesis, the operation moving the [plural] feature allows the insertion of the reflexive exponent that does not violate OCP.

(15) *les* + *lo* → *se los*
 3.+pl.dat 3.-f.-pl.acc reflexive 3.acc.+pl

I will not try to account for the movement of the plural marker by a post-syntactic operation and I will leave the question open. It seems that the template leads the linearization of exponents even after post-syntactic operations; for instance Barceloní (the Catalan variety spoken in Barcelona) seems to adopt a similar solution.

(16) shows the paradigm of Barceloní 3rd person clitics (note that in Barceloní there is the same rule of gender impoverishment already discussed for Spanish), (17) is the list of vocabulary items and (18) shows 3rd person dative + 3rd person accusative clusters.

(16)

	Dative	Accusative	
		-f	+f
-pl	<i>li</i>	<i>l</i>	<i>le</i>
+pl	<i>lzi</i>	<i>lz</i>	<i>lez</i>

(17) *l* ↔ 3
e ↔ +f
z ↔ +pl
i ↔ +loc

(18)

		acc	
		-pl <i>l(e)</i>	+pl <i>l(e)z</i>
dat	-pl <i>li</i>	<i>li</i>	<i>lzi</i>
	+pl <i>lzi</i>	<i>lzi</i>	<i>lzi</i>

We can capture the clusters in (18) through three descriptive generalizations:

1. two person markers do not appear in the same cluster;
2. [+plural] always appears;
3. [+feminine] never appears;

The first generalization can be derived directly from the OCP. Indeed Barceloní does not repair OCP violations impoverishing [number] as Spanish, but it seems to delete the person features as described by the rule (19).

(19) [person] → ∅

Secondly, the Barceloní treatment of [+pl] seems similar to the South American Spanish pattern of floating plural: indeed number is never impoverished, at least it is moved on a different clitic.

Thirdly, I can account for the absence of gender exponents assuming that the operation of gender impoverishment applies to the whole cluster.

I will now illustrate the interaction of these operations analyzing one of the clusters in (18) – repeated here in (20) – as already done for the Spanish example.

(20) *li* + *lez* = *lzi*
 3.-pl.dat 3.+f.+pl.acc

(21) template	pers	gen	num	case	+	pers	gen	num	case
features	3	+f	-pl	dat		3	+f	+pl	acc
impoverishment	3	∅	-pl	dat		3	∅	+pl	acc
insertion	*l OCP!	∅	-pl	dat		<i>l</i>	-	<i>z</i>	-
repair	∅	∅	-pl	dat					
insertion2	-	-	-	<i>i</i>					

In (21) gender is impoverished in both clitics, then the OCP is violated and the [person] feature of the dative clitic is impoverished following the rule (19). Finally, the dative exponent is inserted and linearized in accordance with the template giving the form *lzi*.

I will skip the derivation of all Italian and Barceloní clusters and I sum up in table (22) the repairs accounting for the varieties under analysis. For a wider discussion see Pescarini (2005).

(22)	Repairs	
Spanish	[number] → ∅	
Barceloní	[person] → ∅	[source] → ∅
Italian	[gender] → ∅	

(22) is a plausible picture since it shows a consistent set of operations. There are three impoverishments of phi features and an operation of [source] deletion accounting for the substitution of ablative (= partitive) clitics by locative ones. This set of operations accounts economically for heterogeneous patterns even if it leaves a couple of examples unaccounted for like the Italian one in (23). Indeed a rule transforming a reflexive clitic into a locative one is really problematic. I will try to solve the problem by sketching an alternative approach, which seems more promising because it can account for synthetic clusters without language specific operations.

(23) *si* + *si* = *ci* *si* (**si si*) (Italian)
 reflexive impersonal

3 TOWARDS AN ITEM-BASED APPROACH

(24) describes patterns of substitution: for example (24a) describes the Spanish cluster in (1) where the *l* exponent is substituted by the *s* one. At the same time it shows that when *s* has to be repaired (for example in an impersonal + reflexive cluster violating OCP), it cannot be replaced and the whole cluster is ungrammatical.

(24) (a) *l* → *s* → * (Spanish)
 (b) *s* → *c* → ∅ (Italian)
 (c) *n* → *i* → ∅ (Barceloní)

(24b) describes a similar pattern for Italian: *si* is substituted by *ci* – as in (23) – but when two *ci* co-occur, OCP is repaired deleting one of them as shown in (25b).

(25) (a) *A Roma mi ci porta Antonio* (b) *A Roma ci ei porta Antonio*
 To Rome 1.-pl.acc loc brings Antonio To Rome 1.+pl.acc loc brings Antonio

I suggest that these patterns can be accounted for by a corollary of the Subset principle (11) entailing that when (the insertion of) the most specific item is blocked (for independent reasons), it can be replaced only by a less specific exponent.

Therefore, if *le*, *si* and *n* (the most specific clitics) are replaced by *se*, *ci* and *i* when blocked by OCP, it follows that *se*, *ci* and *i* are less specific than *le*, *si* and *n*.

Secondly, if *se*, *ci* and *i* are never replaced when they violate OCP, it follows that there is no clitic less specific than *se*, *ci* and *i*. Therefore, it means that in Spanish, Italian and Barceloní *se*, *ci* and *i* are *the* less specific clitics. These clitics are ‘elsewhere’ morphemes because their minimal specifications allow them to be *automatically* inserted instead of other clitics when OCP is violated.

For example the *ga* clitic in the Vailate dialect (Lombardia) is a syncretic exponent expressing locative, 1st person plural and 3rd person dative. In my opinion it is an elsewhere clitic because it substitutes the *si* exponent when it violates OCP while the co-occurrence of two *ga* is not allowed. It means that *ga* is the less specific clitic in the paradigm of the Vailate dialect.

The derivation of the Barceloní example is a bit more complex, but if we take *i* as the elsewhere clitic of Barceloní, we can account easily for the cluster in (21). Indeed the *l* exponent is not deleted by an operation of impoverishment – as suggested in section 1 – but just replaced by *i*. Then *i* is deleted because another *i* is still present.

(20) *li* + *lez* = *lzi*
 3.-pl.dat 3.+f.+pl.acc

The hypothesis that every clitic system bears an elsewhere morpheme is supported by the analysis of syncretism. Indeed 3rd person reflexive, locative and partitive exponents have frequently replaced some person clitics (usually 3rd person dative and 1st person plural). The point I want to stress here is that these clitics are used also as elsewhere morphemes in the case of synthetic clusters as shown by the examples in (26) describing the same cluster (3.dat + 3.acc) in

three Italian varieties. These examples show that in Logudorese Sardinian the dative clitic is substituted by the locative, in Campidanese by the reflexive, in the dialect of Castrovillari by the partitive.

- (26) (a) *li* + *lu* = *bi* *lu* (**li lu*) (Logudorese, Sard.)
 (b) *ddi* + *ddu* = *si* *ddu* (**ddi ddu*) (Campidanese, Sard.)
 (c) *li* + *lu* = *nu* (**li lu*) (Castrovillari)

In conclusion I would suggest that there is a deep relation (probably due to underspecification) between the diachronic processes responsible for syncretisms and the synchronic ones responsible for synthetic clusters.

The hypothesis above also presents some counter-examples. In particular the Italian cluster *le lo* should be realized as *ce lo*, because the elsewhere clitic of Italian is *ci*, cf. (24). But the real spell out is *glielo* where *gli* is a contextually motivated allomorph of *l*, sensitive to the feature [-f].

- (27) *le* + *lo* = (**ce*) *glie* + *lo*
 3.+f.dat 3.acc (*elsewhere) 3.-f.dat 3.acc

In my opinion this pattern is consistent with the second part of the Subset principle stating that the most specific vocabulary item (among the underspecified ones) is inserted. Indeed in (27) when *l* violates OCP it is substituted by its allomorph *gli* instead of the elsewhere clitic *ci* because the former is, by definition, more specific than the latter. From this observation we can set different strategies of substitution.

When OCP is violated, a clitic is substituted by:

1. an allomorph, if present, as in (27);
2. an elsewhere clitic, as in (26);
3. \emptyset , if the clitic is an elsewhere, as in (25).

The main point is that the ranking of these strategies is not a stipulation, but it is due to the Subset principle in (11).

4 CONCLUSIONS

Finally, I will conclude with a remark on a question I have left open during the analysis: why is the repaired clitic on the left of the cluster? Firstly, if OCP blocks the insertion of the left clitic, it means that the right one has been already inserted. Therefore, if vocabulary insertion proceeds in accordance with syntactic constituency (Bobaljik 2000), it means that the right clitic is lower than the repaired one.

In this paper I have explored two accounts of synthetic clitic clusters in Romance. The first one is based on language-specific morphological operations that are assumed *ex post*, while the latter is an item-based approach deriving directly from the basic assumption of Distributed Morphology. It seems to me that it provides an elegant model allowing us to predict and test the behaviour of the elsewhere clitic in each variety.

REFERENCES

- Ackema, Peter (2001). 'Colliding Complementizers in Dutch: Another Syntactic OCP Effect'. *Linguistic Inquiry* 32: 4, 717-727.
- Bobaljik, Jonathan David (2000). 'The Ins and Outs of Contextual Allomorphy'. *University of Maryland Working Papers in Linguistics*: 10, 35-71.
- Bonet, Eulalia (1991). *Morphology after syntax: Pronominal clitics in Romance*. MIT: Doctoral dissertation.
- Bonet, Eulalia (1995). 'Feature Structure of Romance Clitics'. *Natural Language and Linguistic Theory*: 13, 607-647.
- Calabrese, Andrea (1994). *Syncretism phenomena in the clitic systems of Italian and Sardinian dialects and the notion of morphological change*. In J.N. Beckman, ed., *Proceedings of NELS 25: 2*. GLSA, University of Massachusetts, Amherst, 151-174.
- Cummins, Sarah & Yves Roberge (1994). 'Romance Inflectional Morphology In and Out of Syntax'. *MITWPL* 22: 53-70.
- Grimshaw, Jane (1997). *The Best Clitic: Constraint Conflict in Morphosyntax*. In L. Haegeman, *Elements of Grammar*. Kluwer.
- Halle, Morris (1990). *An Approach to Morphology*. *Proceedings of NELS*.
- Halle, Morris & Alec Marantz (1993). *Distributed Morphology and the Pieces of Inflection*. In *The View from Building 20*, ed. Kenneth Hale and S. Jay Keyser. MIT Press, Cambridge, 111-176.
- Halle, Morris & Alec Marantz (1994). 'Some key features of Distributed Morphology.' *MITWPL*: 21.
- Harris, James (1991a). 'The Exponence of Gender in Spanish'. *Linguistic Inquiry* 22: 27-62.
- Harris, James (1994). 'The syntax-phonology mapping in Catalan and Spanish clitics.' In *MITWPL*: 21.
- Harris, James (1997). 'Why *n'ho* is pronounced [li] in Barceloní Catalan.' *MITWPL* 30: 451-479.
- Kayne, Richard (2000). *Parameters and Universals*. Oxford University Press.
- Kany, Charles (1945). *American-Spanish syntax*. Chicago: University of Chicago Press.
- Menn, Lise & Brian MacWhinney (1984). 'The Repeated Morph Constraint'. *Language* 60, 519-541.
- Neelman, Ad & Van der Koop, Hans (2004). *Syntactic OCP Effects*. UCL: ms.
- Ortmann, Albert & Alexandra Popescu (2000). 'Haplology Involving Morphologically Bound and Free Elements: Evidence from Romanian'. *Yearbook of Morphology*, 43-70.
- Pescarini, Diego (2005). 'Clitic Clusters and Morphological Repairs. Evidence from Italian, Spanish and Barceloní' in *Studi linguistici e filologici on line*, 3.1: 239-264.
- Simpson, Jane & Meg Withgott (1986). *Pronominal Clitic Clusters and Templates*. In H. Borer, *The Syntax of Pronominal Clitics*. Orlando: Academic Press, 149-174.
- Yip, Moira (1998). *Identity Avoidance in Phonology and Morphology*. In S.G. Lapointe, D.K. Brentari, P.M. Farrel, *Morphology and its Relation to Phonology and Syntax*. Stanford: CSLI, 216-246.

Diego Pescarini

Dipartimento di discipline linguistiche, comunicative e dello spettacolo
via Beato Pellegrino 1
Università di Padova
Padova
35137
Italia

diego.pescarini@unipd.it