

The Best Clitic: Constraint Conflict in Morphosyntax¹

Jane Grimshaw, Rutgers University

in *Elements of Grammar*, Liliane Haegeman ed. 1997, Kluwer Academic Publishers.

Abstract

This paper presents an analysis of object clitic pronouns in Romance; Italian and Spanish in particular. The occurrence of unexpected clitic forms in certain combinations is argued to follow from the effects of a constraint against duplication of clitics. This constraint conflicts with faithfulness constraints which impose faithfulness to input morphosyntactic specifications for case, person, number etc. The ranking of the anti-duplication constraint relative to key faithfulness constraints determines whether a grammar allows duplication, or selects instead a clitic which is less faithful to the input than the clitic which would occur in isolation. The representation of the clitics, as partially specified clusters of morphosyntactic properties, interacts crucially with the faithfulness constraints, which prefer partially specified (less marked) clitic outputs to those which are more fully specified but contradict the input more.

1. Introduction

This paper illustrates the application of Optimality Theory (Prince and Smolensky 1993) to the domain of morphosyntax. It deals with a strictly delineated set of data, concerning the behavior of clitic combinations in the Romance languages, and draws very heavily on research presented in Bonet (1991, 1995).

The phenomenon of interest is the occurrence of what Bonet calls “opaque clitics”: clitics whose form in certain combinations is unexpected given the choice of clitics in isolation. I will argue that these effects can be insightfully characterized under a theory of optimization: the unexpected clitic forms are the best available, given the need to avoid duplicative sequences of clitics.

The situation is one in which there is a conflict between two goals. One goal is to utilize the very best clitic, the one which occurs in isolation. Another goal is to avoid duplication. I will show that variation among languages and dialects can be explained as a choice between different ways of resolving the conflict. One clitic can delete (avoiding duplication), one clitic can change form, or the language can insist on the very best clitic and live with the duplication. These choices are the inevitable result of alternative rankings of the constraints in the system, as expected under optimality theoretic assumptions.

¹ The research reported here was supported by grant SBR-9511891 from the National Science Foundation and by Rutgers University. It has benefited from presentation at Keio University, Stanford University, the University of California at Santa Cruz, and the Third Australian Linguistic Institute at the Australian National University. Special thanks to the Rutgers Optimality Research Group, in particular Eric Bakovic, Ed Keer, Suzanne Preuss, Alan Prince, Melissa Trachtenburg and Vieri Samek-Lodovici for much helpful discussion, and to Eric Bakovic for important insights concerning Spanish.

The chief points of Optimality Theory as presented in Prince and Smolensky (1993) are these. First, all constraints are universal. Second, constraints can be violated in well-formed sentences. A grammar is a ranking of the universal constraints which determines which constraint is satisfied in case of a conflict. The optimal form (that which best satisfies the constraints under the ranking in the grammar) is grammatical, all non-optimal candidates are ungrammatical. An optimal output form for a given input is selected from among the class of competitors in the following way: a form which, for every pairwise competition involving it, best satisfies the highest-ranking constraint on which the competitors conflict, is *optimal*. (Grimshaw to appear). The input in the cases at hand is a set of specifications for morphosyntactic properties: reflexivity, person, number, gender and case. The constraints and their rankings determine what clitic form is the best realization of the input.

1.2 The constraints relevant for opaque clitics

One set of constraints, entirely familiar from the phonological literature, functions to regulate the relationship of the input to the output, imposing a “faithful” relationship between the two. In particular, there are constraints requiring that the input and the output have identical specifications for the morphosyntactic features mentioned above:

FILL FEATURE	Only features in the input can appear in the output
PARSE FEATURE	All features in the input must appear in the output

These constraints are violated whenever an opaque clitic occurs (and elsewhere as we will see).

A further set of constraints targets specific values of these features. *1, *2 and *3 are violated by any form with a first, second or third person specification respectively. I suggest that they form a universal markedness hierarchy, such that *2 dominates *1, and *1 in turn dominates *3 in the grammar of every language, although there is no attempt made here to explore this fully. Similarly there are constraints which penalize case specifications, specifically *DAT for dative and *ACC for accusative, and I suggest that these also form a universal markedness hierarchy in which *DAT dominates *ACC.

The final constraint is the constraint against duplication, which is clearly related to the Obligatory Contour Principle (Leben 1973, McCarthy 1986), which I call “*XX”, and which is violated by combinations of identical elements. This is clearly part of a much more general group of constraints, including the OCP, the double-ing constraint (Ross 1972) constraints on case such as the “double o” constraint of Japanese, and the Case OCP discussed in Mohanan (1994). Assuming that all of these involve heads, the constraint can be stated as ruling out sequences of identical functional heads.

*XX Sequences of identical functional heads are ill-formed

It is the ranking of the faithfulness constraints relative to *XX which supports the system of clitic combinations. When the faithfulness constraints dominate *XX then a faithful parse of the input is the result and the output clitic is the one that would be found in isolation, even though it violates *XX. When *XX dominates a faithfulness constraint, then the clitic will change or disappear in order to satisfy *XX.

The constraint rankings are of two types; the universal markedness hierarchies which were mentioned above, and the language particular rankings.

Universal Markedness Hierarchies

*2 >> *1 >> *3
 *DAT >> *ACC

Some language particular rankings are common to the clitic systems of Italian and Spanish. These, together with the common lexicons discussed shortly, determined a shared core in the clitic systems.

FILL PERS >> PARSE NUM
 FILL R >> PARSE PERS, PARSE NUM, and PARSE CASE

The remaining language particular rankings are the locus of cross-linguistic variation, as mentioned above. I will discuss the alternative rankings and their consequences as they become relevant.

2. The Structure of the clitic system: Italian

While teaching grammars of Italian often describe the clitic system in terms such as “the accusative pronoun clitics, the dative pronoun clitics, the accusative reflexives and the dative reflexives”, this characterization misses some of the fundamental properties of the clitic system. For example, *vi* is the only 2nd person plural form: it is used for reflexive and non-reflexive situations and it is used as a dative and an accusative. *Si* is used for 3rd person reflexives, regardless of their person, number, gender and case. What this suggests is that it is unilluminating at best to characterize each clitic as a fully specified set of morpho-syntactic features: we then have to posit multiple *vi*'s (one for each page of the grammar book), multiple *si*'s and so forth. Alternatively, we can treat clitics as potentially underspecified bundles of morphosyntactic features. In a situation where a clitic with a given specification is called for, an output clitic with exactly that specification will be the best realization, but if there is no such clitic available, an underspecified clitic which best matches the input may suffice. This is where optimality comes into play.

For example, it could be that *si* occurs as the reflexive, third person, plural, feminine, accusative because it is specified for all these features. Or it could be that it is not specified for all of these features but no alternative clitic with a better specification is available. This is the line that will be pursued here.

Once we think about clitics in these terms, it is clear that the different clitics vary in how much morphosyntactic information they encode. *si*, the extreme, encodes none: it is not marked for reflexivity “R”, person, “P”, number, “N”, gender “G”, and case “C”. When a clitic is not marked for a feature, this is indicated by enclosing the abbreviation for the feature in parentheses, hence “(R)” indicates that the clitic is unmarked for reflexivity, i.e. has no reflexivity specification.

Specifically, I will assume the clitic lexicon given in (1).²

(1) Clitic Lexicon of Italian		
si	(R)(P)(N)(G)(C)	<i>self</i>
mi	(R) 1 sg (G)(C)	<i>me, to me, (self)</i>
ti	(R) 2 sg (G)(C)	<i>you, to you, (self)</i>
ci	(R) 1 pl (G)(C)	<i>us, to us, (self)</i>
vi	(R) 2 pl (G)(C)	<i>you, to you, (self)</i>
lo/la	-R (P) sg m/f acc	<i>him, her, it</i>
li/le	-R (P) pl m/f acc	<i>them</i>
gli/le	-R (P) sg m/f dat	<i>to him, her, it</i>

The fault lines in the system involve the R specifications and the P specifications. There is a group of clitics which ranges across reflexive and non-reflexive uses, marked as (R) in (1). A second group is non-reflexive only, -R in (1). With respect to the person dimension, there is a group of clitics which mark person (1st or 2nd) and a group which do not. This latter group will in fact only occur as the output for a 3rd person input, because, as we will see, the clitics which mark person are better outputs for 1st and 2nd person inputs. Finally, there is *si*, which marks nothing. The two fault lines, the R and P specifications, are connected. All and only (R) clitics mark person. Apart from *si*, all and only -R clitics are (P). Other generalizations hold but will not be analyzed here: for example, all and only (R) clitics are (G) and (C), so the (R) clitics don't vary for gender and case and the -R clitics do.

The most interesting clitic in this analysis is *si*, the clitic which is usually described as a reflexive clitic. In this view it is really no such thing. It is a clitic which has no properties. This is in essence the proposal of Bonet (1995), slightly rephrased.

3. Clitics in isolation

The first step is to illustrate the operation of the system in the simple cases. How does it settle on the right lexical choice when a clitic occurs in isolation? We begin with an input for which there is a nearly perfect solution, given the above analysis of the clitic system (I ignore gender here in the interests of simplicity). Assume an input which specifies a non-reflexive 3rd person plural accusative clitic. The actual correct output is *li* or *le* depending on the choice of gender.

² Italian has no clitic form for the plural counterpart of the datives *gli/le*

(2) Input is -R

Input: <[-R 3 pl acc]>

candidates	PARSE R	FILL R	PARSE PERS	FILL PERS	PARSE NUM	FILL NUM	PARSE CASE	FILL CASE
a. si (R)(P)(N)(C)	*		*		*		*	
b. mi (R) 1 sg (C)	*		*	*	*	*	*	
c. ti (R) 2 sg (C)	*		*	*	*	*	*	
d. ci (R) 1 pl (C)	*		*	*			*	
e. vi (R) 2 pl (C)	*		*	*			*	
f. lo/la -R (P) sg acc			*		*	*		
g. li/le -R (P) pl acc			*					
h. gli/le -R (P) sg dat			*		*	*	*	*

Since the winning candidate violates only PARSE PERS, and every other candidate violates this constraint and others, there is no evidence for constraint ranking in this tableau, g. being the winner under every ranking. It is thus impossible to tell from this case which violations are “fatal” and which are incidental. Under these assumptions, all the -R clitics work in the same way: they are all faithful realizations of an input, apart from their failure to parse the person specification. (Later in the paper I will argue that singular and masculine are in fact not marked, hence ultimately the optimal clitic here will violate PARSE GEND as well, if the input is masculine. The constraint system still picks the same candidate as optimal, however, and this should be clear in what follows.)

The (R) clitics do not have a specification for reflexivity. Since every input is specified as either reflexive or nonreflexive, an (R) clitic cannot be an entirely faithful realization of any input. Why, then, are they selected as output forms? Suppose the input is 2nd person reflexive, instead of 3rd. This is illustrated in (3).

(3) Input is R, non-3rd person

Input: <[R 2 pl acc]>

candidates	PARSE R	FILL R	PARSE PERS	FILL PERS	PARSE NUM	FILL NUM	PARSE CASE	FILL CASE
a. si (R)(P)(N)(C)	*		*		*		*	
b. mi (R) 1 sg (C)	*		*	*	*	*	*	
c. ti (R) 2 sg (C)	*				*	*	*	
d. ci (R) 1 pl (C)	*		*	*			*	
e. vi vi (R) 2 pl (C)	*						*	
f. lo/la -R (P) sg acc	*	*	*		*	*		
g. li/le -R (P) pl acc	*	*	*					
h. gli /le -R (P) sg dat	*	*	*		*	*	*	*

All candidates fail PARSE R. The -R clitics will do worse on an R input than the (R) clitics, since the -R specification guarantees failure on FILL R in addition to PARSE R. However, the -R clitics can do better on the case constraints (and the gender constraints, not shown) than the (R) clitics, since the (R) clitics do not parse case. So we must rank FILL R above PARSE CASE (and PARSE GEND) to eliminate the -R group. (In this case, the person specification in the input is such that the -R clitics must fail the person constraints in addition, so it might appear that the ranking of the person constraints might do the job here but this is not true as we will see shortly.) To put the situation in intuitive terms, the system prefers a clitic which does not contradict the R specification of the input over one which faithfully parses the case specification.

For these constraints, this ranking is sufficient to guarantee the choice of *vi* as optimal. It violates only PARSE R, which all the other candidates also violate, and PARSE CASE, which is satisfied only by candidates which violate FILL R. Every other candidate violates at least two additional constraints. However, the constraint *2, to be encountered crucially in the next section, will prefer first and third person clitics to *vi*. In fact, then, it is necessary to rank PARSE PERS over *2, to ensure faithful parsing of the 2 in the input.

Finally, consider *si* in its various incarnations in Italian. As a first example, we can take the form realizing a 3rd person plural reflexive input, as in (4).

(4) Input is R, 3rd person

Input: <[R 3 pl acc]>

candidates	PARSE R	FILL R	PARSE PERS	FILL PERS	PARSE NUM	FILL NUM	PARSE CASE	FILL CASE
a. si (R)(P)(N)(C)	*		*		*		*	
b. mi (R) 1 sg (C)	*		*	*!	*	*	*	
c. ti (R) 2 sg (C)	*		*	*!	*	*	*	
d. ci (R) 1 pl (C)	*		*	*!			*	
e. vi (R) 2 pl (C)	*		*	*!			*	
f. lo/la -R (P) sg acc	*	*	*		*!	*!		
g. li/le -R (P) pl acc	*	*	*					
h. gli/le -R (P) sg dat	*	*	*		*!	*!	*!	*!

In analyzing this tableau, it is helpful to divide it into two parts: the (R) clitics and the -R clitics. Among the (R) clitics, *si* satisfies FILL PERS, which is violated by all the other candidates. *si* violates PARSE NUM, which *ci* and *vi*, the plurals, satisfy. Hence *si* will be victorious if FILL PERS >> PARSE NUM. Among the non-reflexive clitics, *li* or *le* is the winner. All other candidates share the R and PERS faithfulness violations and violate other constraints in addition.

Now we can do a reduced comparison, looking at the optimal candidate from each of the two parts of the previous tableau:

(5) Reduced comparison: the best (R) clitic versus the best -R clitic

Input: <[R 3 pl acc]>

candidates	PARSE R	FILL R	PARSE PERS	FILL PERS	PARSE NUM	FILL NUM	PARSE CASE	FILL CASE
a. si si (R)(P)(N)(C)	*		*		*		*	
g. li/le -R (P) pl acc	*	*!	*					

This comparison reveals a crucial ranking: FILL R >> PARSE NUM, and PARSE CASE. (Evidence for FILL R >> PARSE CASE has already been given.) For a reflexive input, the grammar of Italian chooses a clitic that is not irreflexive in form, even though it does not encode the input properties of number and case. Faithfulness to the reflexivity of the input, in so far as this is possible in the system, is given priority over faithfulness to its number, case (and also gender, not discussed here).

Note that both constraint rankings are crucial. The ranking of FILL R >> PARSE NUM, PARSE CASE successfully chooses the (R) clitic over all -Rs, but it will not choose *si* from among the (R) clitics. FILL PERS >> PARSE NUM is needed for that.

The appearance of *si* as the impersonal subject is illustrated in (6):

- (6) Lo **si** sveglia
 3rd-acc impers. wakes-up
 “one wakes him up”

Must we stipulate that *si* is chosen as the impersonal subject? Given the analysis proposed here, which follows Bonet’s insight, *si* is the least marked clitic, having no specifications. Thus if the grammar of Italian stipulates that the impersonal is a clitic, the choice of *si* is predictable. Suppose, for example, that the input for an impersonal is just [3 pl]. (Impersonal subjects trigger plural participial agreement, Cinque (1988).).

- (7) The impersonal is the least marked clitic

Input: <[3 pl]>

candidates	PARSE R	FILL R	PARSE PERS	FILL PERS	PARSE NUM	FILL NUM
a. si (R)(P)(N)(C)			*		*	
b. mi (R) 1 sg (C)			*	*!	*	*!
c. ti (R) 2 sg (C)			*	*!	*	*!
d. ci (R) 1 pl (C)			*	*!		
e. vi (R) 2 pl (C)			*	*!		
f. lo/la -R (P) sg acc		*!	*		*	*
g. li/le -R (P) pl acc		*!	*			
h. gli/le -R (P) sg dat		*!	*			

The clitic *si* violates PARSE NUM, but FILL R dominates this constraint, so all -R candidates are eliminated (*si* violates only PARSE PERS in addition, and every other candidate violates this constraint too.) The singular (R) clitics *ti*, *mi* violate everything that *si* does plus additional constraints, so they are eliminated. The plural (R) clitics satisfy PARSE NUM, but violate FILL PERS. Since FILL PERS >> PARSE NUM, *si* beats *ci* and *vi*. So the constraints and rankings which chose *si* in the previous case, choose *si* now, and it is not an accident at all that same clitic is used for both.

The observation that the impersonal subject clitic is the same as the reflexive clitic has been a puzzling one: what do they have in common? (See Cinque 1988 for a detailed discussion of the

uses of *si* in Italian.) This line of reasoning suggests that what they have in common is not some elusive reflexivity or other syntactic/semantic property, but the simple absence of morphosyntactic detail. (A reasonable further target, then, would be the extension of this solution to the so-called “inherent reflexives”, where the clitic (*si* with a third person subject, *mi* with a first person subject etc.) does not, at least not obviously, correspond to an argument at all. If the grammar of Italian specifies that the verb form involves a clitic, it should be possible to predict the choice of clitic from the constraints given here.)

So far *si* has the property that it violates every PARSE constraint, since it parses nothing in the input. It also satisfies every FILL constraint: since it has no specification for any of the morphosyntactic features, it cannot have a specification which contradicts that of the input. However, it cannot be the case that *si* parses nothing at all. If *si* parses nothing at all then it is necessarily equivalent to the null candidate. Why, then, is it preferred to realize the clitic at all, rather than deleting it? The null candidate, like the *si* candidate, would violate the PARSE constraints and not the FILL constraints. The problem is illustrated in (8), for the input analyzed above in (4)-(5).

(8) *si* versus null candidate, without constraint PARSE CL

Input: <[R 3 pl acc]>

candidates	PARSE R	FILL R	PARSE PERS	FILL PERS	PARSE NUM	FILL NUM
a. si (R)(P)(N)(C)	*		*		*	
b. (R)(P)(N)(C)	*		*		*	

The answer must be that there is a further constraint which prefers parsing of any kind over no parsing at all. For present purposes let us call this constraint PARSE CLITIC, though it is perhaps more general in form. Furthermore, the proper analysis of *si* must be that it does parse the feature Cl, even though it parses nothing else. Now the constraint will eliminate the null candidate, as illustrated in the table in (9), which compares the null candidate with *si* for the same [R 3 pl acc] input.

(9) *si/se* versus null candidate, with constraint PARSE CL

Input: <[Cl R 3 pl acc]>

Candidates	PARSE CL	PARSE R	FILL R	PARS PERS	FILL PERS	PARSE NUM	FILL NUM
a. si (R)(P)(N)(C)		*		*		**	
b. (R)(P)(N)(C)	*!	*		*		**	

On any ranking among these constraints, PARSE CL will select *si* over the null candidate.

Subsequently we will see that alternative rankings of PARSE CL control the choice between deletion of a clitic and its realization. Until they become important again, however, I will omit Cl from inputs, and the PARSE CL constraint from tableaux.

In the analysis developed here, a crucial role is played by the ranking of $\text{FILL PERS} \gg \text{PARSE NUM}$. There is another possibility here and throughout, in which si is selected by a markedness constraint $*\text{PERS}$, a constraint which dislikes person specifications. When PARSE PERS is violated, $*\text{PERS}$ will choose a form which has no person specification over one which has the wrong one. In a system where $*\text{PERS} \gg \text{PARSE PERS}$, there would be no clitics marked for person, and the same point holds for all features. I will not pursue this possibility here, postponing it for a further study which addresses the issue of how the lexicon of clitics is to be properly understood.

In this paper, I take the clitic lexicon as a given, and show how the constraint system derives constancy of forms in isolation and variation in combination. However, one of the interesting points of Optimality Theory is that it makes it possible to derive grammatical inventories themselves from rankings of constraints (Prince and Smolensky (1993), Grimshaw (to appear), Grimshaw and Samek-Lodovici (to appear)). A constraint such as *PERS finds its natural place in such an analysis. Although I do not develop this point here, it should be noted that even in the present paper, the constraint ranking is playing a critical role in determining the lexicon of clitics that is actually realized in these languages.

Tableau (10) illustrates this point for a relatively straightforward example. Recall that the choice of *vi* as the optimal candidate for [R 2 pl acc] input requires the ranking FILL R >> PARSE CASE. This was shown in (3) above. This is because the *vi* clitic fails to parse the case in the input, and the -R clitics can successfully do this. Hence we must count the FILL R violation as dominating the effects of PARSE CASE. What happens under other rankings? The answer is that we can derive a system with no second person pronouns. (10) shows the fatal violation of PARSE CASE that *vi* suffers.

(10) PARSE CASE >> FILL R, PARSE PERS, FILL PERS

Input: <[R 2 pl acc]>

If PARSE CASE dominates FILL R and the person constraints, then the optimal candidate will be the -R which successfully parses the case, even though it counterspecifies reflexivity and fails both person constraints. Such a grammar will have no second person pronoun. (The same choice is optimal for an input with a -R specification, where the g. candidates satisfy FILL R in addition to the constraints they satisfy in (10).) This is an illustration of the interplay between constraint rankings and lexical inventories, documented in Prince and Smolensky (1993).

Returning to the main theme, the rankings established so far are these:

FILL PERS >> PARSE NUM

FILL R >> PARSE NUM, and PARSE CASE

These are constant across all the Romance dialects and languages discussed here. We will see shortly that other rankings, those involving *XX in particular, show significant variation.

4. Clitic combinations: *XX

The central hypothesis of this paper is that clitic combinations are regulated by a violable constraint like the Obligatory Contour Principle, which prohibits clitic duplication. The ranking of this constraint relative to the faithfulness constraints determines what happens in clitic combinations. When *XX dominates a faithfulness constraint, the faithfulness constraint will be violated and the form which occurs will be the one that best fits the input given the constraint rankings. It may even be the case that only one of the input clitic specifications is realized in the output, the other deleting. When *XX is dominated by the relevant faithfulness constraint(s), the *XX violation will be ignored and the clitics which occur will be exactly those that occur in isolation.

The data in (11) (from Bonet 1995) illustrates the alternation found in Italian between *si* and *ci*. The clitic *si* is the best output for the impersonal input, as we saw above in (7), and it is also the best output for a 3rd person reflexive input (as shown above in (4-5)). Both of these clitics occur in combination with other clitics, as (11a,b) show (although *si* is replaced with *se* in clitic combinations, a substitution which I will not analyze here.) When we would expect to find two occurrences of *si*, however, or perhaps *si* plus *se*, this is not possible. Instead, we find the combination *ci* plus *si*. This use of *ci* is an example of what Bonet (1995) calls an “opaque clitic”: *ci* is generally the first person plural clitic but in this situation it is not encoding first person.

- C. **Ci si lava** impersonal plus reflexive (*si si)
 ci si washes
 “one washes oneself”

Let us assume for simplicity that *ci* in the output corresponds to the impersonal subject input (which is plural in Italian as noted above), although the same prediction is made with respect to choice of forms if *ci* corresponds to the reflexive. To realize the impersonal input with *si* violates *XX. To realize it with *ci* results in an incorrect person specification, hence the ranking of FILL PERS below *XX is crucial.

(12) *si si > ci si* (Tableau ignores case and gender, also candidates with dative -R clitics)

Input: <[3 pl] [R 3 pl]>

Candidates	*XX	PARSE R	FILL R	PARS PERS	FILL PERS	PARSE NUM	FILL NUM
a. si+si (R)(P)(N)+(R)(P)(N)	*	*		**		**	
b. li/le +si -R (P) pl + (R)(P)(N)		*	*!	**		*	
c. lo/la+si -R (P) sg + (R)(P)(N)		*	*!	**		**	*
d. ci ci+si (R) 1 pl +(R)(P)(N)		*		**	*	*	
e. mi+si (R) 1 sg + (R)(P)(N)		*		**	*	**!	!*
f. vi+si (R) 2 pl + (R)(P)(N)		*		**	*	*	
g. ti+si (R) 2 sg + (R)(P)(N)		*		**	*	**!	!*

The singular clitics in c. e. and g. are the same as the corresponding plurals, except that they have a violation of FILL NUM plus an extra violation of PARSE NUM, so the singular candidates can be eliminated. Since all candidates fail PARSE R once and PARSE PERS twice, we can eliminate these constraints from consideration. The resulting greatly reduced tableau is in (13).

(13) *si si > ci si* (Tableau ignores case and gender, also candidates with dative -R clitics)

Input: <[3 pl] [R 3 pl]>

Candidates	*XX	FILL R	FILL PERS	PARSE NUM	FILL NUM
a. si+si (R)(P)(N)+(R)(P)(N)	*			**	
b. li/le +si -R (P) pl + (R)(P)(N)		*!		*	
d. ci ci+si (R) 1 pl +(R)(P)(N)			*	*	
f. vi+si (R) 2 pl + (R)(P)(N)			*	*	

Recall that *ci* is not the best clitic for the impersonal input when it is in isolation. This is because it counter-specifies the person of the input, and since FILL PERS >>PARSE NUM, the fact that it successfully parses the number in the input, which *si* does not, is irrelevant. In order for *ci* to be selected here, then, there is a crucial ranking between *XX and FILL PERS. The ranking selects *ci* with its FILL PERS violation over *si*. We also see in (13) evidence for a further ranking. The non-reflexive plural clitics can realize the input with no FILL PERS violation, because they have no person specification, while *ci* is specified as 1. The non-reflexive clitics do, however, have a -R specification and thus induce a violation of FILL R, hence provided that FILL R >> FILL PERS the non-reflexive candidates will be eliminated. This assumes that the input is not -R, of course, as does the original analysis of impersonal *si* in (7).

Finally, the markedness constraint *2, which universally dominates *1, selects *ci* over *vi*. (Note that the existence of *1 is inferred only, there is no direct evidence for such a constraint. If it does exist, then it must be ranked below *XX, in order that the violation of *1 which *ci* incurs not result in incorrect choice of *si* as optimal.) The crucial language particular rankings involved in the *si/ci* alternation are thus those in (14).

- (14) Crucial Rankings: *XX >> FILL PERS, FILL R >> FILL PERS, *XX >> *1

5. Re-ranking to get the *si si* dialect

Bonet (1995) notes that there is a dialect of Italian (Conegliano), illustrated in (16), which does not show the *si/ci* alternation. Here instead, *si* appears twice.

- (15) Standard Italian
Ci si lava
ci si washes
“one washes oneself”
- (16) Conegliano
Si si lava
si si washes
“one washes oneself”

Bonet observes that this is quite puzzling because it undermines any attempt to derive the *si/ci* alternation from a universal constraint against like occurring with like. If such a constraint existed, it would rule out the existence of Conegliano. If such a constraint does not exist, how do we explain the *si/ci* alternation? Under an optimality theoretic account, this dilemma is resolved. It is both the case that the general constraint exists and the case that grammatical sentences can violate it. This is what will arise if the *XX constraint is dominated by a crucial faithfulness constraint. In particular, if FILL PERS >> *XX, as in (17), the grammar will select the *si si* output, since it satisfies FILL PERS.³

³ Alternatively, assuming the existence of *1 (see above), the ranking of *2, *1 >> *XX will give the *si si* dialect, regardless of the ranking of *XX and FILL PERS.

(17) Re-ranking of *XX and FILL PERS

Input: <[3 pl] [R 3 pl]>

Candidates	PARSE R	FILL R	PARS PERS	FILL PERS	*XX	PARSE NUM	FILL NUM	*2	*1
a. si+si (R)(P)(N)+(R)(P)(N)	*		**		*	**			
d. ci+si (R) 1 pl +(R)(P)(N)	*		**	*!		*			*

Thus the existence of the Conegliano dialect is predicted: it is simply the result of re-ranking a faithfulness constraint with *XX.

6. Spanish *se* ellipsis

As noted earlier, there is more than one way to resolve the problem posed by clitic duplication. One is to change one of the clitics into another form as in Italian, another is to delete one. Bakovic (p.c.) points out that Spanish opts for the latter in at least some dialects: the generality of this phenomenon remains to be determined.

- (18) a. *Se se lava
“one washes oneself”
b. Se lava

In the earlier discussion of the success of *si* and the success of the null candidate at parsing input specifications, we saw that since *si* parses none of the person, number, gender and reflexivity of the input, it is preferred over the null candidate only by virtue of its success in satisfying the constraint PARSE CL which is satisfied if the output is any clitic, regardless of how unsatisfactory a parse the clitic may offer. The “reduction” of *se se* to just *se* results if the *XX constraint dominates PARSE CL, provided that FILL PERS >> PARSE CL also. This is illustrated in the tableaux below, in which *XX and FILL PERS both dominate PARSE CL.⁴

⁴ A slight complication is that while Italian impersonal subjects seem to be plural (Cinque 1988, 536-7), their Spanish counterparts seem to be singular. I will show the Spanish input as [3 sg], and hence consider *me* in Spanish as the counterpart to *ci* in Italian, in assessing potential as the output for the impersonal.

(19) $se \ se > se$ (Tableau radically simplified)

Input: <[Cl 3sg] [Cl R 3 sg]>

Candidates	*X X	PARSE R	FILL R	PARS PERS	FILL PERS	PARSE CL	PARSE NUM	FILL NUM
a. se+se Cl (R)(P)(N)+Cl (R)(P)(N)	*!	*		**			**	
b. me+se Cl (R) 1 sg+.Cl (R)(P)(N)		*		**	*!		*	
c. me se Cl (R)(P)(N)		*		**		*	**	

Since PARSE R, FILL R, PARSE PERS and FILL NUM do not distinguish among the key candidates, we can eliminate them all and simplify the tableau as follows.

(20) $se \ se > se$

Input: <[Cl 3sg] [Cl R 3 sg]>

Candidates	*XX	FILL PERS	PARSE CL	PARSE NUM
a. se+se Cl (R)(P)(N) + Cl (R)(P)(N)	*!			**
b. me+se Cl (R) 1 sg + Cl (R)(P)(N)		*!		*
c. me se Cl (R)(P)(N)			*	**

Candidate a. is eliminated in favor of c. if *XX >> PARSE CL: the two candidates are identical on all other constraints. This leaves candidates b. and c. The reduced candidate c. does worse than b. on two constraints: PARSE CL and PARSE NUM, and better only on FILL PERS, so the following rankings are required in addition:

FILL PERS >> PARSE CL, PARSE NUM

We now need to briefly reconsider what was said above concerning Italian and the Conegliano dialect, taking the effects of PARSE CL into account. Clearly, since all the key candidates perform the same with respect to PARSE R, FILL R, PARSE PERS and FILL NUM, we can continue to eliminate them from the tableaux. We need only consider possible alternative rankings among the remaining constraints.

The next tableau shows the ranking which yields the Conegliano facts: here PARSE CL and FILL PERS both dominate *XX.

(21) *si si > si si**Input: <[Cl 3 pl] [Cl R 3 sg]>*

Candidates	PARSE CL	FILL PERS	*XX	PARSE NUM
a. si+si Cl (R)(P)(N) + Cl (R)(P)(N)			*	**
b ci+si Cl (R) 1 sg + Cl (R)(P)(N)		*!		*
c. si Cl (R)(P)(N)	*!			**

Note the crucial role played in (21) by the ranking of FILL PERS over *XX. If PARSE CL >> *XX then a form with two clitics will always result when the input has two units, but whether the two clitics are identical as in (21) or distinct as in (22) depends on the ranking of *XX and FILL PERS. The choice of *si si* over *ci si* depends also on the previously discussed ranking of FILL PERS over PARSE NUM, which is violated one more time in the optimal candidate than in the *ci si* candidate.

(22) *si si > ci si**Input: <[Cl 3 pl] [Cl R 3 sg]>*

Candidates	*XX	PARSE CL	FILL PERS	PARSE NUM
a. si+si Cl (R)(P)(N) + Cl (R)(P)(N)	*!			**
b. si Cl (R)(P)(N)		*!		**
c. ci+si Cl (R) 1 pl + Cl (R)(P)(N)			*	*

When *XX dominates PARSE CL and PARSE CL dominates FILL PERS, the *ci si* pattern is the outcome, since the counterspecification of person is less serious in its consequences than the duplication of clitics or failure to parse a clitic. The same output is selected if PARSE CL dominates *XX and PARSE CL >> FILL PERS. In other words, it results from every system where PARSE CL >> FILL PERS, and *XX >> FILL PERS, regardless of the ranking of *XX and PARSE CL.

The general result is as follows. The duplication of *si* is a point of conflict in the constraint system. On the one hand, faithfulness requires realization as *si* for both inputs, but on the other hand *XX prohibits this outcome. What we have seen is that a variety of resolutions are possible, depending on the ranking of a few key constraints: *XX, PARSE CL, FILL PERS and PARSE NUM. Depending on how they are ranked, the solution will be to live with the duplication, to delete a clitic, or to replace one of the offending clitics with a less faithful alternative.

7. Spanish “spurious *se*”.

The data in (23), again taken from Bonet (1995), illustrate the phenomenon known as “spurious *se*” in Spanish (Perlmutter 1971). When a third person accusative and a third person dative combine, the dative is replaced by *se*.

- (23) a. E1 premio, **lo** dieron a Pedro ayer.
the prize 3rd-acc gave(3rd-pl) to Pedro yesterday
- b. A Pedro, **le** dieron el premio ayer.
to Pedro 3rd-dat gave(3rd-pl) the prize yesterday
- c. A Pedro, el premio, **se lo** dieron ayer. (***le lo** ***lo le**)
to Pedro the prize se 3rd-acc gave(3rd-pl) yesterday
“they gave the prize to Pedro yesterday”

An analysis of spurious *se* must explain why the expected output does not occur, why it is the dative that is replaced and not the accusative, and why *se* appears rather than some other clitic.

The general structure of the solution here is clear: the *le lo* combination violates some constraint which dominates the relevant faithfulness constraint and *se* appears in a not unexpected fashion. Assuming for now that it is *XX that *le lo* violates, we have the situation in (24).

- (24) *XX induces *le* > *se*.

Input: <[-R 3 sg dat] [-R 3 sg acc]> (case constraints omitted from consideration)

Candidates	*XX	PARSE R	FILL R	PARSE PERS	FILL PERS	PARSE NUM	FILL NUM
a. <i>le + lo -R (P) sg dat + -R (P) sg acc</i>	*!						
b. le <i>se + lo (R)(P)(N)(C) + -R (P) sg acc</i>		*		*		*	
c. <i>me + lo (R) 1 sg (C) + -R (P) sg acc</i>		*		*	*		
d. <i>te + lo (R) 2 sg (C) + -R (P) sg acc</i>		*		*	*		
e. <i>ci/vi +lo (R) 1/2 pl (C) + -R (P) sg acc</i>		*		*	*	*	*

However, (24) also reveals that it is rather unclear why *le + lo* should violate *XX. They are not phonologically identical, unlike the cases we have seen so far. Moreover, although they are morphologically similar in that they are both descriptively third person, in the analysis given here they are unmarked for person, and represent the residual case. They are, of course, both -R clitics, and perhaps this suffices to trigger *XX. What makes this uncertain is the question of what other kinds of morphological identity might count for the constraint, and why. (Note that they are not morphologically identical, because one is dative and one accusative, moreover the spurious *se* effect is found when a singular combines with a plural, so identity of number is not relevant.)

These issues remain unresolved at this point. There are three ways to think about the nature of the constraint violation in the *le lo* combination. One is that it is really phonological, and the identity of onsets is sufficient to violate *XX. The second is that it is really morphological, and that the occurrence of two -R clitics violates *XX, or perhaps, if we revise the analysis of 3rd person, that the duplication of 3 is in violation of the constraint. The third possibility is that the source of the violation is a little different, and concerns the constraints on clitic order, which have sometimes been analyzed in terms of a “template” or output constraint, following the original proposal by Perlmutter (1971). Bonet proposes such a solution for certain opaque forms in Catalan. According to this view, the realization of clitics must conform to principles determining what count as good clitic sequences. For Spanish, Perlmutter (1971, 45) proposes the following condition: *se II I III*. According to this constraint, then, there is only one third person position (“III”), so one might hypothesize that the clitic combination *le lo* (and indeed some of the others discussed here) violates the templatic requirement and hence cannot be realized. Since there is only one position for a 3rd person in the template, a clitic which is not third person must “replace” a third person form in combinations.

While this idea has some initial appeal, its ultimate fate depends on the theory of these templatic restrictions. On further reflection it is hard to see how the templatic view really differs from saying that two (descriptively) 3rd person clitics cannot combine. In other words, it seems highly likely that the template requirements themselves are the result of the system of constraints at work in the selection of clitics. Recognizing that this issue needs to be resolved, I will continue to assume that *XX is at work here. Clearly, though, its status as the key constraint here is leading into questions beyond those we can address.

Returning to the tableau in (24), given the clitic inventory, all -R candidates violate *XX, for one of the reasons discussed above. The best of these candidates must be *le* which best parses the input, hence of the -R clitics only *le* is shown in the tableau. The ranking *XX >> PARSE R eliminates all the -R clitics.

For familiar reasons, *se* is the best choice of the remaining clitics. Of the candidates which satisfy *XX b. satisfies FILL PERS, which the others violate. But b. violates PARSE NUM, which c. and d. satisfy. Therefore FILL PERS >> PARSE NUM yields *se* as optimal.

Crucial rankings	$\begin{array}{l} \text{*XX} >> \text{PARSE R} \\ \text{FILL PERS} >> \text{PARSE NUM} \end{array}$
------------------	-----------------------------------------------------------------------------------------------------

Since PARSE CASE is violated in the optimal candidate, but satisfied in the candidate which violates *XX, we can also deduce that *XX >> PARSE CASE; however, the case constraints are not included in the tableau.

The next question is why *le*, rather than *lo*, is replaced by *se*. Tableau (25) shows the alternatives available if *le* is retained. The best candidate in this set is *le* plus *se*, candidate b, given the rankings just mentioned.

(25) Why keep *lo* and not *le*?

Input: <[-R 3 sg dat] [-R 3 sg acc]>

Candidates	*XX	PARSE R	FILL R	PARSE PERS	FILL PERS	PARSE NUM	FILL NUM
a. le + lo -R (P) sg dat + -R (P) sg acc	*!						
b. le + se -R (P) sg dat + (R)(P)(N)(C)		*		*		*	
c. le + me -R (P) sg dat + (R)1sg(C)		*		*	*	*!	
d. le + te -R (P) sg dat + (R) 2 sg (C)		*		*	*	*!	
e. le + vi/ci -R (P) pl dat + (R) 1/2 pl (C)		*		*	*	*!	*

When we compare *le* plus *se* with *lo* plus *se*, it becomes clear that they are identical except with respect to the case on the retained item, accusative in one instance, dative in the other. Since each offers a faithful parse of the input case specification, no faithfulness constraint relating to case can distinguish them. Hence *le* + *se* and *se* + *lo* are equally (un)successful parses for the input, and a markedness constraint must be at work. A constraint against dative (*DAT), universally ranked above a constraint against accusative if such exists (here *ACC), will have the right consequences.

(26) Best of candidates which retain *lo*, and best of candidates which retain *le*:

Input: <[-R 3 sg dat] [-R 3 sg acc]>

Candidates	*XX	PARSE R	FILL R	PARSE PERS	FILL PERS	PARSE NUM	FILL NUM	*DAT	*ACC
a. le se + lo (R)(P)(N)(C) + -R (P) sg acc			*	*		*			*
b. le + se -R (P) sg dat + (R)(P)(N)(C)			*	*		*		*!	

In this analysis, then, we can predict that *le* *lo* will be replaced by *se* *lo*. The only puzzle concerns the precise understanding of *XX in this context, raising the question of whether some other constraint will prove to lie behind the effect.

8. Reranking to derive *le* *lo*.

Given the preceding discussion, it comes as no surprise that not every Romance language shows the spurious *se* effect. If Parse R >> *XX, then the transparent clitics will surface, as they do in Italian (27), where *gli* plus *lo* surfaces as the complex *glielo*:

(27) Glielo venderà
You will sell it to him/her

This ranking picks the candidate which parses the -R specification of the input, and the best such candidate is the one which appears in isolation.

- (28) *XX below Parse R, no spurious *se*
 input <[-R 3 sg dat] [-R 3 sg acc]> (case constraints omitted from consideration)

Candidates	PARSE R	*XX	FILL R	PARSE PERS	FILL PERS	PARSE NUM	FILL NUM
a. gli + lo -R (P) sg dat + -R (P) sg acc		*					
b. se + lo (R)(P)(N)(C) + -R (P) sg acc	*!			*		*	
c. mi + lo (R) 1 sg (C) + -R (P) sg acc	*!			*	*		
d. ti + lo (R) 2 sg (C) + -R (P) sg acc	*!			*	*		
e. ci/vi +lo (R) 1/2 pl (C) + -R (P) sg acc	*!			*	*	*	*

9. Floating Number

Bonet shows that in some dialects of American Spanish, the accusative clitic takes on the number associated with the dative argument, under spurious *se* conditions. (29a) is Iberian Spanish, (29b) the American Spanish dialect version.⁵

- (29) a. El libro, a ellos, ¿quién **se lo** prestó?
the book to them who se 3rd-acc lent(3rd)
 “who lent the book to them?”
 b. El libro, a ellos, ¿quién **se los** presto?

The general idea here is a simple one. In a spurious *se* configuration, the input specification is not well parsed. In particular the number of the dative argument in the input is unparsed since *se* is unspecified for number. In (29b) the system selects the option of expressing the plurality of the argument “replaced” by *se*, by expressing it on the other argument. Thus (29b) involves suppressing the singular character of the direct object in order to express the plural character of the indirect object.

The conclusion we have to draw is that there is an asymmetry, not addressed here previously, between singular and plural. This motivates a further step in the analysis of the basic clitic systems, which is illustrated in (30)-(31). The input contains two number specifications, one for the accusative clitic and one for the dative. In the case where the number specifications of the two clitics differ, if the number of the dative clitic is realized on the accusative, it follows that the number of the accusative clitic is not realized. If, on the other hand, the number of the accusative is realized, the number of the dative will not be. Thus in either case, under assumptions made so far, there will be a violation of PARSE NUM. This is illustrated in (30). Since the realization of the number of the

⁵ Kany (1951, 109-112 cites many examples of both floating number and floating gender (analyzed below), commenting that “grammarians and purists inveigh against this solecism”.

dative clitic on the accusative clitic will also violate **FILL NUM**, there is no way that the candidate with the number transferred can ever be optimal. (In the greatly reduced tableaux to be used here, only the constraints on which the candidates differ are included, and the constraint violations induced by *se*, those involved in the spurious *se* configuration, are not represented on the tableaux.)

- (30) Why sg must be unmarked

Input: <[-R 3 pl masc dat] [-R 3 sg masc acc]>

Candidates which satisfy *XX	PARSE NUM	FILL NUM
se + lo (R)(P)(N)(G)(C) + -R (P) sg masc acc	*(pl)	
se + los (R)(P)(N)(G)(C) + -R (P) pl masc acc	*(sg)	*!

The conclusion we must draw is that the singular must be unmarked, and the plural marked. More precisely, it must be the case that the representation of the singular clitics has no number specification, so that *lo* is properly represented as in (31a) instead of (31b), as we have assumed so far.

- (31) a. [-R (P) (N) masc acc]
 b. [-R (P) sg masc acc]

Under this representation, we still get the right result for candidates in isolation, as we can see in (32), which shows the outcome for a singular input.

- (32) Selecting the right candidate in isolation: singular

Input: <[-R 3 sg masc acc]>

Candidates	PARSE NUM	FILL NUM
los -R (P) pl masc acc	*	*!
☞ lo -R (P) (N) masc acc	*	

When the input is plural, *los* will be selected:

- (33) Selecting the right candidate in isolation: plural

Input: <[-R 3 pl masc acc]>

Candidates	PARSE NUM	FILL NUM
los -R (P) pl masc acc		
lo -R (P) (N) masc acc	*!	

Returning to the problem of floating number, and utilizing the asymmetry between singular and plural, the situation can now be summarized slightly differently.. When the accusative clitic is singular, its number specification is never parsed, because the available clitics are unmarked for number. There is thus no additional cost with respect to parsing the number of this clitic if it is realized as a plural: the input singular is unparsed whether the clitic used in *los* or *lo..* In isolation, the option of realizing a singular input as *los* is prevented by FILL NUM, which penalizes the realization of the plural specification on the accusative clitic, as we saw in (32). In the floating number situation, therefore, there is a conflict between the goal of parsing the input number specification of the dative clitic, and improperly attaching plural specification on the singular input accusative argument. To put it in constraint terms, PARSE NUM and FILL NUM are in conflict in this case. If PARSE NUM >> FILL NUM, the number will float.

(34) Floating number if PARSE NUM >> FILL NUM

Input: <[-R 3 pl masc dat] [-R 3 sg masc acc]>

Candidates	*XX	PARSE NUM	FILL NUM
se + lo (R)(P)(N)(G)(C) + -R(P) (N) masc acc		**!	
☒ se + los (R)(P)(N)(G)(C) + -R (P) masc acc		*	*

Realizing the accusative input with a plural clitic satisfies PARSE NUM for the plural specification, but violates FILL NUM on the accusative. There is a crucial assumption behind this proposal, which needs to be refined. The assumption is that PARSE NUM is satisfied in the optimal candidate, despite the fact that the number specification is parsed in the wrong clitic. This is clearly the result that we want: it is better to parse the number in the wrong place than not to parse it at all. It is assumed here that a FILL NUM violation ensues when the accusative is realized as a plural, because the FILL constraint looks at the input for that clitic, rather than the input as a whole. In order to explicate this more precisely we need to draw upon the results of recent phonological research on correspondence, and I will turn to this shortly.

10. Floating Gender

The more general prediction of this analysis is that for any input specification, its realization will float if PARSE F >> FILL F. Bonet (1995) discusses dialects of American Spanish, citing colloquial Mexican and Uruguayan, in which *se la* is the preferred output when the input dative is feminine, even when the input accusative is masculine. She cites the data in (35b), which contrasts with the Iberian Spanish (35a).

- (35) a. Si ella me quiere comprar el caballo, yo **se lo** venderé.
if she 1st-dat wants buy the horse I se 3rd-acc will-sell (1st)
 “If she wants to buy my horse, I will sell it to her”
- b. Si elle [sic] me quiere comprar el caballo, yo **se la** venderé.
if she 1st-dat wants buy the horse I se 3rd-acc-fem will-sell(1st)
 “If she wants to buy my horse, I will sell it to her”

For reasons entirely parallel to those discussed for singular number, this situation requires that masculine gender is unmarked and, more precisely, that there are no clitics marked as masculine. The correct representation of *lo* is as in (36a) rather than (36b).

- (36) a. [-R (P) (N) (G) acc]
 b. [-R (P) (N) masc acc]

I will not go through the details of how the correct candidate is selected in isolation, as the logic is exactly the same as for number. As was the case with number, in this analysis the “masculine” clitics do not parse the gender specification of the input, and hence always violate PARSE GEND even when the clitics occur in isolation. Hence *se* plus *lo* in (35a) violates PARSE GEND twice, while *se* plus *la* in (35b) violates the constraint only once. Provided, then, that PARSE GEND >> FILL GEND, gender will float as number does.⁶

- (37) Floating gender if PARSE GEND >> FILL GEND

Input: <[-R 3 sg fem dat] [-R 3 sg masc acc]>

Candidates	PARSE GEND	FILL GEND
se + lo (R)(P)(N)(G)(C) + -R (P) (N) (G) acc	**!	
☒ se + la (R)(P)(N)(G)(C) + -R (P) (N) fem acc	*	*

The same assumptions hold here concerning what count as violations of the PARSE and FILL constraints as above.

11. Reranking for dialects with no floating features:

Not surprisingly, in view of the previous discussion, the opposite ranking of the PARSE and FILL constraints yields dialects in which there is no floating. So where FILL F >> PARSE F, F will not float. Tableau (38) illustrates this for gender.

- (38) No floating gender if FILL GEND >> PARSE GEND

input <[-R 3 sg fem dat] [-R 3 sg masc acc]>

Candidates	FILL GEND	PARSE GEND
☒ se + lo (R)(P)(N)(G)(C) + -R (P) (N) (G) acc		**
se + la (R)(P)(N)(G)(C) + -R (P) (N) fem acc	*!	*

⁶ It is not clear why gender floats only with spurious *se*, and not also when the dative clitic is a first or second person, all of which are just like *se* in failing to encode gender.

Finally, if the PARSE constraints for both gender and number dominate the FILL constraints, then both number and gender will float. The data in (39), again from Bonet (1995), illustrate this: (39a) is Iberian Spanish, (39b) illustrates the relevant American dialects.

- (39) a. Si ellas me quieren comprar el caballo, yo **se lo** venderé.
if they(fem) 1st-dat want(3rd-pl) buy the horse I se 3rd-acc will-sell(1st)
 “if they want to buy my horse, I will sell it to them”
- b. Si ellas me quieren comprar el caballo, yo **se las** venderé.
if they(fem) 1st-dat want(3rd-pl) buy the horse I se 3rd-acc-fem-pl will-sell(1st)
 “if they want to buy my horse, I will sell it to them”

In this case the optimal candidate violates both FILL NUM and FILL GEND, but satisfies both PARSE NUM and PARSE GEND.

- (40) Gender and Number float together

Input: <[-R 3pl fem dat] [-R 3 sg masc acc] >

Candidates which satisfy *XX	PARSE GEND	FILL GEND	PARSE NUM	FILL NUM
se + lo (R)(P)(N)(G)(C) + -R (P) (N) (G) acc	**!		**	
se + la (R)(P)(N)(G)(C) + -R (P) (N) fem acc	*	*	**!	
se+los (R)(P)(N)(G)(C) + -R (P) pl (G) acc	**!		*	
se+las (R)(P)(N)(G)(C) + -R (P) pl fem acc	*	. *	*	*

I should point out that, in this analysis, the floating of gender and the floating of number are independent, so we expect to find dialects in which neither float, dialects in which both float, and dialects in which only one of the two floats.

As I mentioned above, there is a crucial assumption behind this proposal, which needs to be refined. The assumption is that the PARSE constraints are satisfied when the features float, even though the features are parsed in the “wrong” clitic, but the FILL constraints are violated in this case; this requires assuming that the PARSE constraints look at the input as a whole while the FILL constraints look at the input for individual clitics. It is hard to see why such an asymmetry should be expected. We can explicate the situation rather differently, if we draw upon the results of recent phonological research on correspondence, presented in McCarthy and Prince (1995). In their discussion of reduplication and identity requirements they develop three constraint families. Considering only their role in constraining the relationships between inputs and outputs, the MAX family requires that every segment of the input has a correspondent in the output, the DEP family requires that every segment in the output has a correspondent in the input and the IDENT(F) family requires that correspondent segments are identical in feature F. The key point is that MAX and DEP

only require that each piece of the input has a correspondent in the output, or vice versa, whereas the IDENT(F) constraints are sensitive to the location of the correspondent. Let us take the position that the number feature in the input is in correspondence with the number feature of the output and so forth, and that the set of feature specifications in the input is in correspondence with the set of feature specifications in the output. We can now apply this revised analysis to the case of floating number, for illustrative purposes, showing, as always, only the violations that result from the selection of the second clitic, and not those triggered by *se*.

(41) Floating number if PARSE NUM >> FILL NUM

Input: <[-R 3 pl masc dat] [-R 3 sg masc acc]>

Candidates		MAX NUM	DEP NUM	IDENT NUM
se + lo (R)(P)(N)(G)(C) + -R (P) (N) (G) acc	*!			*
☒ se + los (R)(P)(N)(G)(C) + -R (P) pl (G) acc				**

MAX holds for features and for the clitics as collections of features. In particular, it is violated for every feature in the input which does not have a correspondent in the output, so there are 5 violations overall for *se* in both candidates, three for *lo* in the first candidate and two for *los* in the second candidate. Crucially, MAX NUM is violated in *se lo* but not in *se los*. DEP NUM (and more generally DEP for all the output specifications) is satisfied in both candidates, since no feature in the output lacks a corresponding feature in the input, assuming crucially that the “pl” feature of the output is in correspondence with the “sg” specification in the input, ie. that different values for the same feature, here number, are in correspondence with each other. Finally, IDENT NUM is violated by *se* in both candidates, since the correspondents are not identical with respect to number (taking the correspondents to be the collections of features between square brackets in the input and the collections of features between plus signs in the outputs). One correspondent, the input clitic specification, has number specified (among other features) while the other correspondent, the output clitic specification, has no number. IDENT NUM is also violated by the combination “sg in input plus plural in output” for the *los* candidate, hence this candidate has two violations of the constraint. What is going on, then, is that the fact that the plural feature is realized at all satisfies MAX, but IDENT NUM is violated in reflection of the fact that the plural specification is not in the correspondent clitic. IDENT NUM is playing the role here that FILL NUM played previously, penalizing the realization of the feature on the wrong clitic, while MAX NUM is playing the role in which PARSE NUM was previously cast, with the rather odd consequences noted earlier. The critical ranking, then, is MAX NUM >> IDENT NUM, and the other ranking will give dialects with no floating features.

That constraints developed in phonological research should provide an improved solution for the problems of clitic choice should hardly be surprising by now. In particular, both reduplication systems and the clitic systems under study here are domains in which notions of faithfulness and identity play a crucial role. In any event, it seems that separating the faithfulness constraints into these three types, as opposed to the previous two types, offers the conceptual clarification needed for a solution of floating number and gender.

13. Conclusions

There are a number of important issues which are not addressed at all in this paper, but require further investigation. One is the question already raised, of whether the optimality theoretic account can extend in an interesting way to the generalizations concerning clitic order, usually taken to reflect the operation of some kind of surface structure template as first discussed in Perlmutter (1971). The second issue, more directly challenging to the theory, is how to explain why certain clitic combinations are simply impossible. This is the case with, for example, Italian combinations like *mi mi*, where one is a dative and one an accusative, and for others discussed in Wanner (1977). Here there is no change in the clitic choice which results in grammaticality, and the only possible form is one in which one of the arguments is expressed as a phrase and not as a clitic. Optimality Theory offers two lines of attack on this problem. The ill-formed combination may be impossible because in fact there is a better output, one which better satisfies the constraints. It is possible, for example, that a phrasal expression competes with a clitic expression for status as the optimal output for a given input. Alternatively, the combination may be impossible because the null candidate is the optimal candidate in this case, hence no actual candidate survives. Both of these lines of analysis need to be pursued.

The choice of clitics in these Romance systems is highly systematic, as Bonet (1995) has already shown. The argument of this paper is that this systematicity can best be explicated in terms of the notion of optimization: the clitic used is the best one available in the circumstances. The “circumstances” are provided by the clitic inventory of the language and the rankings of the grammatical constraints, especially the ranking of faithfulness constraints relative to *XX. In this way, the constraints of Universal Grammar determine the actual set of words which appears in the sentence which realizes a given input.

Cross-linguistic variation here is limited to the effects of constraint rankings: the same basic clitic system exists in each case. The constraints themselves are stated over a set of independently needed morphosyntactic features, such as “person”, “reflexive”, and simply require faithful parsing of these features. The same constraints regulate the clitic alternations studied here as regulate the more mundane case of clitic choice where no alternations occur. Finally, we see the importance of the optimality theoretic notion of “minimal violation”. While it is perfectly possible to select a clitic which violates a constraint, indeed many constraints, such violation must be compelled by a higher ranking constraint. Violation is not free. Hence we see clitic alternations under pressure from *XX, but clitics in isolation are not affected by this constraint, and hence appear in a standard form with no alternations.

References

- Bonet, i Alsina, M. Eulàlia, 1991. *Morphology after Syntax: Pronominal Clitics in Romance*, doctoral dissertation, Massachusetts Institute of Technology.
- Bonet, Eulàlia, 1995. "Feature Structure of Romance Clitics", *Natural Language and Linguistic Theory* 13, 607-647.
- Cinque, Guglielmo, 1988. "On *si* Constructions and the Theory of *Arb*", *Linguistic Inquiry* 19, 521-582.
- Grimshaw, Jane, 1997. "Projection, Heads and Optimality", *Linguistic Inquiry* 28, 373-422.
- Grimshaw, Jane and Vieri Samek-Lodovici, (to appear). "Optimal Subjects and Subject Universals", in Barbosa, Fox, Hagstrom, McGinnis and Pesetsky eds. *Is the Best Good Enough?* MITWPL/ MIT Press.
- Leben, Will, 1973. *Suprasegmental Phonology*, Doctoral dissertation, Massachusetts Institute of Technology.
- McCarthy, John, 1986. "OCP Effects: Gemination and antigemination", *Linguistic Inquiry* 17, 207-263.
- McCarthy, John and Alan Prince, 1995. "Faithfulness and Reduplicative Identity", in J. Beckman, L.W. Dickey and S. Urbanczyk, eds., *University of Massachusetts Occasional Papers in Linguistics* 18, 249-384. Rutgers Optimality Archive, ROA-60.
- Mohanan, Tara, 1994. "Case OCP: A Constraint on Word Order in Hindi" in Miriam Butt, Tracy Holloway King and Gillian Ramchand Eds. *Theoretical Perspectives on Word Order in South Asian Languages*, CSLI, 185-216.
- Perlmutter, David, 1971. *Deep and Surface Structure Constraints in Syntax*, Holt, Rinehart and Winston.
- Prince, Alan and Paul Smolensky, 1993. *OptimalityTheory: Constraint interaction in generative grammar*. RuCCS Technical Report #2. Rutgers University Center for Cognitive Science, Piscataway NJ.
- Ross, John Robert, 1972. "Doubl-ing", *Linguistic Inquiry* 3, 61-86
- Wanner, Dieter, 1977. "On the Order of Clitics in Italian", *Lingua* 43, 101-128.