Variable Clitic Sequences in Nonstandard French: Feature Geometry or Optimality?

David Heap  
UWO, djheap@uwo.ca

Svetlana Kaminskaia  
skaminsk@uwo.ca

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/pwpl/vol7/iss3/9
For more information, please contact libraryrepository@pobox.upenn.edu.
Variable Clitic Sequences in Nonstandard French: Feature Geometry or Optimality?
1 Introduction

Recurring questions involving sequences of pronominal clitics in Romance languages include those listed in (1):

(1) a. How do pronominal clitics get placed in sequence within a formal grammar?
   b. How can we represent intragrammatical variation in clitic sequences?
   c. How can we account for opaque clitic sequences?

While (1a) is a classic problem (dating from Perlmutter 1971) within the generative tradition, and (1c) has received a fair amount of attention recently (see Bonet 1991, 1994, 1995 among others), (1b) has not. This paper discusses the first two questions as applied to nonstandard varieties of French. We argue that both the sequencing of pronominal clitics and the intragrammatical variation in their sequencing can best be captured by assuming that the internal morphological structure of clitics is determined by a Feature Geometry, which acts (among other things) as a limit on the power of constraints linearizing clitics within Optimality Theory (as proposed by Grimshaw 1997, 1999).

Previous studies have described the order of the pronominal clitics in Romance languages on the basis of either syntactic movement (Bastida 1976, Pearce 1991, Laenzlinger 1993, Uriagereka 1995) or morphological templates (Perlmutter 1971; Harris 1994, 1996; Nadasdi 1995; Miller & Sag 1997). For the most part, these proposed analyses only deal with standard...
Romance varieties (but see for example Morin 1979, 1981), and neither approach has been able to provide a satisfactory formal solution for the problem of clitic sequencing, even within an (idealized) invariant grammar. And with respect to clitic sequences which vary within a given grammar, neither optional syntactic movements nor alternating morphological templates can actually motivate variability in clitic ordering, much less constrain it. Crucially, since the orders of clitic sequences are largely, but not completely, fixed in some varieties (Hetzron 1977, Todoli 1995, Wanner 1996, Heap 1998), how can a formal grammar allow just the amount of variation that is attested, and no more?

In a Standard French sentence, pronominal clitics are organized as in (2), where clitics on the left precede clitics on the right, and clitics within a column mutually exclude each other.²

(2) a. Preverbal position:
   se > me > le > lui > y > en
   te la leur
   nous les vous

b. Postverbal position:
   le > lui > y > en
   la leur
   les moi (m’)
   toi (t’)
   nous vous soi

In this paper we will only be concerned with enclitic (postverbal) orderings of the type in (2b). This order, which is fixed in Standard French, varies both between and within related nonstandard varieties. It has been proposed elsewhere that this type of intragrammatical variation can be captured within Optimality Theory using unranked and floating constraints (Reynolds & Nagy 1994, Anttilla 1997, Nagy & Reynolds 1997, Heap 1998, Nagy & Heap 1998, Auger & Steele 1999, Auger 2000). But OT on its own is too powerful to account for variable clitic ordering in a nonstipulative fashion. Constraints of the ALIGN family (Anderson 1996a,b; Grimshaw

² Sequences of more than three clitics seem to be unattested in langue d’oil. Sequences with ethical datives are not considered in this paper. ✸
1998, 2000) can represent any order (and therefore, predict little or nothing). What complicates the situation is that such analyses depend on a Clitic Lexicon that is based on a matrix of unordered binary features, and this dramatically over-generates the inventory of possible clitics (Heap 2000b).

It is therefore desirable to formulate more restrictive constraints and representations such that the orders (be they fixed or variable) of clitics within sequences are a direct consequence of their internal morphological structures. Under this proposed approach (Heap 1999, 2000a), unranked or floating constraints within OT permit intragrammatical variation, but a theory of Feature Geometry (Bonet 1991, 1995; Harley & Ritter 1998; Béjar 1999, 2000) is required in order to impose some limits on its generative power.

2 Feature Geometry

General theoretical justifications for the use of a Feature Geometry approach include the following (Harley & Ritter 1998: 3):

(3) a. Cross-linguistic variation and paradigm-internal gaps and syncretisms are constrained by the hierarchical organization of features in the universal geometry.

b. The interpretation of sub-trees of the geometry may be relativized so that the language-specific interpretation of a given feature will depend in part upon the contrasts available within the feature system of that language.

The geometry proposed for French clitics is given in (4). It is a modified and simplified version of Harley & Ritter (1998).

(4)
Here, the root node is simply CL (for clitic, rather than RE for Referring Expression, i.e. all pronouns and agreement markers); additional number and class distinctions not needed here are also omitted. In such a geometry, organizing nodes are in Bold, and features, which are privative, are [bracketed] terminal nodes.

2.1 Markedness

The geometry proposed above directly represents the degree of specification (markedness) as the amount of internal structure of a clitic. This approach in turn reflects some assumptions about markedness adopted from Rice (1999):

- Markedness correlates with structure: more complex structures are more marked, less complex structures are less marked.
- Markedness is not a single dimension: which features within a class function as marked depends on a number of factors (including the inventory and the amount of structure required to distinguish the contrasts within it).
- Markedness can vary positionally: different features may be unmarked in different positions.

An explicit theory of markedness (such as this one) can also help predict the loci of intra- and intergrammatical variation: following Anttila (1997), we assume that variation occurs where a given grammar is underspecified. As Wanner states with respect to the “precedence conditions” which he proposes to account for Spanish clitic sequences, such “local effects [are] in principle variable according to regions, periods, people, and styles” (1996:33, our translation). Our hypothesis is that the position of a clitic in a string also follows from its markedness (or specification), and that where the clitic inventory is underspecified, variation in sequencing can occur.

2.2 Applications to Clitic String Orderings

In Standard Spanish (Heap 1998, 1999, 2000b) the fixed clitic sequences as well as the variable sequences attested in some varieties (such as Murcian), can be captured by just one constraint (Heap 1998, 2000b) which orders clitics in a “crescendo effect” (Harris 1996). This constraint, as stated in (5), arranges clitics from the morphologically least specified to most specified:

(5) **Least Leafy to the Left (LLL) Constraint**: Clitics are linearized according to their degree of morphological specification, from least
specified (at left) to most specified (at right).

The question which then arises is: Can the LLL constraint, along with the Feature Geometry proposed for Standard and nonstandard Spanish varieties, be applied to clitic strings in nonstandard northern French (langue d'oil) varieties? In order to attempt to answer this question, we analyzed clitic sequences from 113 points (i.e. different geographical locations) in the northern half of the Atlas Linguistique de la France (Gilliéron & Edmont 1902-1910) (henceforth referred to as the ALF), the broadest survey available of nonstandard forms from this region. In Tables 1 and 2 we present summaries of the clitic sequences obtained from the ALF maps #410 and #411: these maps present the forms provided at each survey point as the local dialect (or patois) translation of the Standard French enclitic sequences dis-le-moi and dis-le-lui, respectively. While it would obviously have been very useful to compare these with other forms (combinations of different clitics, as well as strings in different syntactic environments), these two maps represent the only ALF data which provide sequences of pronominal clitics.

2.2.1 Variants of dis-le-moi in Nonstandard French

Table 1 (opposite) presents the data from Map #410 (dis-le-moi). For the present discussion, we will not be referring to the data in parts 4 and 6 of this table, since doubled pronouns (i.e. forms such as moi...moi, me...moi, l...lle, etc.) fall beyond the scope of this analysis. ALF survey point numbers are not given for types 1 and 2, as they are too numerous to list here; the geolinguistic distribution of these forms is left for future study. Note that example forms are given in IPA transcription rather than the original Gilliéron-Rousselot alphabet.

The most frequent order is Type 1, moi-le or me-le (attested in 58 locations). This type of form follows directly from the LLL constraint as formulated in (5), given that under the geometry proposed in (4), the less specified me precedes the more specified le:

\[
\begin{align*}
\text{CL} & \quad \text{CL} \\
\text{PARTICIPANT} & \quad \text{INDIVIDUATION} \\
[speaker] & \quad \text{CLASS} \\
me & \quad le
\end{align*}
\]
<table>
<thead>
<tr>
<th>Types of sequences</th>
<th>Number of occurrences</th>
<th>Examples (forms in 3, 4, 5, 6 only)</th>
<th>ALF point numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. moi - le order:</td>
<td>(total: 58)</td>
<td>[dimel(l)e] type 39</td>
<td>widespread</td>
</tr>
<tr>
<td>a) /dimel(l)e/ type</td>
<td>39</td>
<td>/dimel(l)e/ type 19</td>
<td></td>
</tr>
<tr>
<td>b) /dimwalls/ type</td>
<td>19</td>
<td>/dimwalls/ type 14</td>
<td></td>
</tr>
<tr>
<td>2. le - moi order:</td>
<td>(total: 33)</td>
<td>[dil:lami] type 19</td>
<td>widespread</td>
</tr>
<tr>
<td>a) /dilami/ type</td>
<td>19</td>
<td>/dilami/ type 14</td>
<td></td>
</tr>
<tr>
<td>b) /dilamwa/ type</td>
<td>14</td>
<td>/dilamwa/ type 14</td>
<td></td>
</tr>
<tr>
<td>3. Strongpronoun:</td>
<td>(total: 10)</td>
<td>[dil:la] 13</td>
<td></td>
</tr>
<tr>
<td>a) instead of moi:</td>
<td>dis-le à moi</td>
<td>[dil:la] 13</td>
<td></td>
</tr>
<tr>
<td>b) instead of le:</td>
<td>dis-moi ça</td>
<td>[dil:la] 13</td>
<td></td>
</tr>
<tr>
<td>4. moi doubled by</td>
<td>(total: 5)</td>
<td>[dimole] 63</td>
<td></td>
</tr>
<tr>
<td>a strong pronoun:</td>
<td>dis-moi-le à moi</td>
<td>[dimole] 63</td>
<td></td>
</tr>
<tr>
<td>a) order moi - le:</td>
<td>dis-moi-le à moi</td>
<td>[dimole] 63</td>
<td></td>
</tr>
<tr>
<td>b) order le - moi:</td>
<td>dis-le-moi à moi</td>
<td>[dimole] 63</td>
<td></td>
</tr>
<tr>
<td>5. Two variants</td>
<td>(total: 6)</td>
<td>[dimell] ~ [dimel] 167</td>
<td></td>
</tr>
<tr>
<td>co-occurring at</td>
<td></td>
<td>[dimell] ~ [dimel] 167</td>
<td></td>
</tr>
<tr>
<td>one survey point</td>
<td></td>
<td>[dimell] ~ [dimel] 167</td>
<td></td>
</tr>
<tr>
<td>6. le doubled</td>
<td>(total: 1)</td>
<td>[dilmelle] 272</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>113</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Summary of ALF Map #410 "dis-le-moi"

The next most frequent order is Type 2, le-moi or le-me (attested at 33 ALF points). Here we can assume an underspecified representation for le (see (7)), since the masculine accusative form has the default gender and case. Thus in grammars of this type the two clitics have the same amount of morphological specification:
As a result, *le* can either precede or follow *me/moi*, since in the event of equal featural specification, the **LLL** constraint correctly predicts that both orders can surface within a single grammar. Because the **ALF** survey protocol favored the listing of a single response for each survey point, cases of equal featural specification such as in (7) would most likely have been recorded as either Type 1 or Type 2; in a handful of cases, however, both clitic orderings are attested as co-occurring in a single grammar (see the six Type 5 survey points in Table 1).

These facts provide support for Antila's (1997) claim that variation will occur where grammars underspecify the outcome: there are two possible clitic orderings which are compatible with the **LLL** constraint, and both surface in these cases. Note that this type of underspecification is likely more common than the six points documented here might seem to suggest, since many such cases could have been recorded on **ALF** Map #410 as having just one or the other of the possible orders.

Crucially, this variation occurs with both *me* and *moi* and is thus independent of the form (stressed or unstressed) of the 1psg pronoun (contra the claim of Laezlinger (1994) and Terzi (1999) that this sort of variation in order is due to the special prosodic status of the form *moi*). The fact that we find both orders (*me le* and *le me*) co-occurring in one grammar shows that this variation has to do with the internal morphological structure of these clitics, rather than with their prosodic nature. Thus, the (variably underspecified) morphological structure of clitics determines their sequencing: full specification of *le* makes it follow *me*, its underspecified representation allows it (optionally) to precede *me* (and in a few cases, to both precede and follow it).

---

3 As John Charles Smith, Daniel Ezra Johnson and Jacques Lamarche, among others, have pointed out, this account has the (probably undesirable) consequence of predicting that a third person clitic specified as feminine and/or plural would necessarily follow a first person clitic: we should find *me la, me les* but not *la me, les me*. Unfortunately, the ALF does not provide any data with which to test such predictions.
2.2.2 Variants of *dis-le-lui* in Nonstandard French

Table 2 summarizes data from *ALF* Map #411 (*dis-le-lui*). Once again, *ALF* survey point numbers are not given for the most widespread types (Types 1, 2, and 3).

<table>
<thead>
<tr>
<th>Types of sequences</th>
<th>Number of occurrences</th>
<th>Examples (forms in 4 and 5 only)</th>
<th><em>ALF</em> point numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Type /di((C)C)V/</strong>&lt;br&gt;forms: /dili/ /diji/ /ditli/ /dizi/ /dizzi/ /dile/ /dil0/</td>
<td>(total: 72)</td>
<td></td>
<td>wide-spread</td>
</tr>
<tr>
<td>/dili/</td>
<td>38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/diji/</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ditli/</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/dizi/</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/dizzi/</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/dile/</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/dil0/</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>2. Forms without le:</strong>&lt;br&gt; <em>dis-lui</em></td>
<td>(total: 8)</td>
<td></td>
<td>wide-spread</td>
</tr>
<tr>
<td><strong>3. Forms where both clitics can be distinguished:</strong>&lt;br&gt;a) order le - lui</td>
<td>(total: 12)</td>
<td></td>
<td>wide-spread</td>
</tr>
<tr>
<td>b) order lui - le</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>4. Forms with a strong pronoun</strong>&lt;br&gt;a) instead of le:&lt;br&gt; <em>dis-lui ça</em></td>
<td>(total: 10)</td>
<td>[dilqis:i:] [dijis:i:]</td>
<td>400 451</td>
</tr>
<tr>
<td>b) instead of lui:&lt;br&gt; <em>dis-le à lui</em></td>
<td>2</td>
<td>[dijias:i:] [deilie]</td>
<td>13 32</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>['dilielij'] ['diloei:]</td>
<td>44 78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>['dizialij']</td>
<td>146</td>
</tr>
<tr>
<td></td>
<td></td>
<td>['diliajli:]</td>
<td>286</td>
</tr>
<tr>
<td></td>
<td></td>
<td>['diliajli:]</td>
<td>287</td>
</tr>
<tr>
<td></td>
<td></td>
<td>['dili:lqi:]</td>
<td>476</td>
</tr>
<tr>
<td><strong>5. Two variants co-occurring at one survey point</strong></td>
<td>(total: 7)</td>
<td>[dilqi:] ~ [dizi] [dili:] ~ [dizii:]</td>
<td>232 239</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[dili:] ~ [dilile] [diji:] ~ [dili:]</td>
<td>262 270</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[dili:] ~ [dilelali] [di:] ~ [diji]</td>
<td>284 343</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[dili:] ~ [dili:le]</td>
<td>465</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>109</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Summary of *ALF* Map #411 "*dis-le-lui*"
In these data, the question of clitic ordering is overshadowed in most cases by the presence of opaque strings (all Type 1 examples in Table 2), that is, surface outputs in which the two underlying clitics are not individually recognizable, such as [dili], [dizi], etc. This tendency to avoid the expression of two third person clitics in a row is attributable to a morphological instantiation of the Obligatory Contour Principle (OCP), a (general, and likely universal) constraint which bans sequences of identical elements (here, morphological rather than phonological features).

The effect of the OCP can either be to merge the two third person clitics into one surface form, to eliminate one of them, or to push one of them out of the clitic string to surface as a stressed pronoun (see Type 4 in Table 2). At this point we do not have an analysis (morphological or phonological) for the opaque single clitic forms of Type 1 (at 72 occurrences, the most frequent type in Table 2), but Types 2 and 4 can be understood as the result of the relative rankings of two constraints, the general fidelity constraint PARSE, restated in (8), and a markedness constraint, AVOID STRONG PRONOUN, following (Bonet, 1994), as in (9):

(8) **PARSE**: All input morphemes must be present in the output.

(9) **AVOID STRONG PRONOUN (ASP)**: Spellout as a clitic is preferred to spellout as a strong pronoun.

Different rankings of the OCP, PARSE, and ASP constraints account for variation in the output sequences: the dominance of PARSE over ASP, and of OCP over both of them, selects the output with a strong pronoun. As shown in (10), candidates (a) and (b) violate the highest ranked OCP constraint, while candidates (c) and (d) violate the second highest constraint, PARSE. Both optimal candidates in (e) only violate the lowest constraint, ASP:

(10) **OCP >> PARSE >> ASP**, input: *le + lui*

<table>
<thead>
<tr>
<th>candidates</th>
<th>OCP</th>
<th>PARSE</th>
<th>ASP</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) le lui</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) lui le</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c) le</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>d) lui</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>e) le ... à lui</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e) lui ... ça</td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

Alternatively, ranking ASP over PARSE, and OCP over both, gives preference to the single pronoun output, as shown in (11). Here, sequences of
two third person clitics, as in (11a) and (11b), violate the highest ranked OCP constraint. Strong pronouns, such as both forms in (11e), violate ASP. Thus, the optimal outputs are the candidates in (11c) and (11d), the latter of which corresponds to Type 2 in Table 2. Further research is required to determine the nature of the constraint(s) which would motivate the preference for one of the two optimal candidates in a given grammar, such as those in (10e), and those in (11c) and (11d).

(11) OCP >> ASP >> PARSE, input: le + lui

<table>
<thead>
<tr>
<th>candidates</th>
<th>OCP</th>
<th>ASP</th>
<th>PARSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) le lui</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) lui le</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c) ☐ le</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>d) ☐ lui</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>e) le ... à lui</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>lui ... ça</td>
<td></td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

Since the OCP disfavors output containing two third person clitics, the question of ordering two clitics only arises in a small minority of cases (the 12 Type 3 points in Table 2). In the small number of cases where both clitics do occur on map #411, the most frequent tendency is to respect the LLL constraint by ordering the least specified case (accusative) before the more specified case (dative), cf. (6) above. This less marked Type 3a order le lui (8 cases attested in Table 2) is illustrated below:

(12) CL
    ┌── INDIVIDUATION ─┐
    │                │
    │                │
    │                │
    │                │
    │                │
    └── CLASS ──────┘
    ┌── GENDER ─┐
    │ CASE ─┐
    le     |

For the relatively marked ordering lui le (4 cases attested in Table 2), we can assume complete specification, as in (13): if we specify masculine gender on the accusative pronoun le, then both clitics have the same degree of morphological specification. As with (7) above, clitics with identical
degrees of morphological specification can be ordered either *le lui* or *lui le*, and the latter order appears as Type 3b in Table 2.

(13)

\[
\begin{array}{c}
\text{CL} \\
\text{INDIVIDUATION} \\
\text{CLASS} \\
\text{GENDER} \quad \text{CASE} \\
(\text{masculine}) \quad \text{le} \\
\end{array}
\begin{array}{c}
\text{CL} \\
\text{INDIVIDUATION} \\
\text{CLASS} \\
\text{GENDER} \quad \text{CASE} \\
(\text{dative}) \quad \text{lui} \\
\end{array}
\]

In this section, we have seen that a sequence composed of two third person clitics can, due to the OCP constraint, surface as one clitic accompanied by a strong pronoun or as just a single clitic. Where OCP is violated and both third person clitics surface together, they can be ordered by the LLL constraint, with either fully specified or underspecified morphological structures. The varieties with the opaque single clitics, and those where two variants co-occur, (Types 1 and 5 respectively in Table 2) await further study.

3 Conclusions

In this article we have considered a range of variable clitic sequences obtained from data in northern French dialects which correspond to the Standard French imperatives *dis-le-moi* and *dis-le-lui*. Our analysis suggests that Optimality Theory can be used to represent the type of variation attested in nonstandard clitic sequences (both within and across grammars), but only if OT is itself constrained by input in which the amount of morphological specification of a clitic (i.e. its markedness) corresponds directly to its degree of internal structure. A Feature Geometry (which can be variably underspecified) provides just such a morphological structure, capable of constraining the number of possible orders within clitic strings, while permitting a certain flexibility with respect to featural (under-) specification, which in turn seems to correspond to the range of variation attested in the forms studied. This confirms the idea that grammars appear to be sensitive to an item’s overall degree of morphological complexity (Béjar 1999, 2000; Béjar & Currie Hall 2000). The LLL or “crescendo” constraint accounts for many of the overall clitic sequencing facts, while other constraints, including
the (morphological) OCP, are necessary in order to account for certain clitic outputs. Further research into variability in clitic ordering, including sequences with different clitics in different morphosyntactic contexts, is required in order to refine our hypotheses regarding the complex interaction between internal morphological structure and constraints on outputs.

References


Department of French
University of Western Ontario
London, ON N6A 5B8
djheap@uwo.ca
skaminsk@uwo.ca