

MORPHOSYNTACTIC THEORY NEEDS IBERIAN LINGUISTIC GEOGRAPHY (AND VICE-VERSA)

David Heap

University of Western Ontario (Canada)

djheap@uwo.ca

Abstract

Spanish clitic sequences, long thought to be strictly ordered, turn out to show variation in the order of first and second person pronouns with the reflexive clitic *se*. This variation shows up in dialect data spanning the 20th century, from the *Atlas lingüístico de la Península Ibérica* (www.alpi.ca) and other linguistic atlas surveys as well as semi-directed interviews with older rural speakers from the *Corpus oral y sonoro del español rural*. The analysis of nonstandard clitic sequences reveals an implicational asymmetry, whereby nonstandard orders with first person singular *me* imply those with second person *te* but nonstandard sequences with *te* can occur alone, without those involving *me*. Such a person asymmetry in turn suggests that accounts of clitic ordering, whether they rely on templates, syntactic movement or ALIGN constraints, must pay attention to the internal morphological structure of pronominal paradigms — something we only notice if morphosyntactic analyses and geolinguistics mutually inform one another.

Key Words

linguistic atlas; Iberian Peninsula; ALPI; nonstandard data; clitic sequences

This paper explores how one long-standing debate in the morpho-syntactic description of Spanish, with implications for morpho-syntactic theory generally, can benefit from data which come from outside the tradition of formal linguistic theory. Specifically, I show how data from dialect atlas surveys can shed genuinely new light on otherwise recalcitrant problems, and that the two areas —linguistic theory and linguistic geography —cannot afford to ignore each other.

A classic issue of morpho-syntactic description in the generative tradition and one of the first where Spanish data figured prominently and crucially in debate on linguistic

theory is the notion of Surface Structure Constraint, as originally proposed more than 35 years ago:

1. Output condition on clitic pronouns:

| | | | |
|-----------|----|---|-----|
| <i>se</i> | II | I | III |
|-----------|----|---|-----|

(Perlmutter 1971:45)

This Surface Structure Constraint succinctly states the possible grammatical output orders of clitics in terms of grammatical persons, rather than in terms of grammatical function: independently of whether they are direct or indirect objects, ‘inherent’ reflexives or ethical datives, clitics in Spanish must appear in this order (*se* followed by second person, followed by first person, followed by third person pronouns). The notation in 1 describes the possible orders of clitics in Spanish, and Perlmutter goes on to make a more general claim that in languages like Spanish:

2. Clitics are *strictly ordered*. (Perlmutter 1971:46, original italics).

All of which seems descriptively uncontroversial: no matter how they may be derived syntactically, sequences such as those in 3a which adhere to 1 are grammatical, while those that do not adhere to the template in 1, such as those in 3b, are not:

3. a. *se nos, se lo, te la, me las*
 b. **nos se, *lo se, *la te, *las me*

Most of the debate around Surface Structure Conditions (or *templates* as they came to be called) in Spanish and in other languages over the intervening decades has been over the exact theoretical status of the notation as in 1 (see among others Dinnsen 1972, Szabo 1974, Wanner 1977, 1994): while a template can adequately *describe* the possible output sequences of clitics in a language such as Spanish, the same notation could equally be used to describe a large number of possible but unattested templates corresponding to nonexistent grammars. There has never been any principled theory constraining what

forms templates such as 1 can or cannot take. The template notation radically over-predicts because it is a completely unconstrained (i.e. overly powerful) device which can be added to morpho-syntactic derivations without any explicit motivation other than to described the attested facts. In other words, clitic templates and related formalisms represent an ad hoc fix which get us out of an empirical bind but without shedding any explanatory light at all on the phenomena described.

As alternatives to the template notation in 1, other proposals attempted to show that clitics could arrive in these fixed orders via some kind of movement from argument positions to positions adjoined to the verb. Without going into details of each proposal (see Heap and Roberge 2001 for an overview), these movement analyses typically tend to fail for the same empirical reasons that originally motivated Perlmutter's (1971) proposal: the internal orders of clitic sequences are conditioned by grammatical person, not by grammatical function or argument position. Any attempt to determine the fixed order of clitics which depends only on the syntactic positions from which they might have moved collide with this inconvenient fact: orders like those described in 1 are independent of the syntactic or argumental origins of the clitics. While various solutions have been proposed which distinguish for example third person clitics from first and second person clitics based on some positional difference in syntactic heads (e.g. Emonds 1975, Bastida 1976, Herchensohn 1980, Pearce 1991, Laenzlinger 1993, Uriagereka 1995), these accounts are equally ad hoc in that they posit syntactic movements which exist specifically and exclusively to describe clitic movement into the attested positions and orders, and which do not correspond to any other syntactic movements in the grammar. Because they rely on otherwise unmotivated movements, such accounts do not provide any generalizable insights into the nature of clitic orderings or syntactic movement.

On the other hand, accounts which rely on 'base-generation' of clitics in their surface positions (rather than syntactic movement) still run into the problem of how to constrain their order when there are two or more clitics. For example, Bonet (1991, 1994, 1995) uses a template which makes reference to the internal morphological structure of clitics, and Harris (1994, 1996) proposes 'precedence conditions' which undertake to

constrain the ordering relationships between different (groups of) clitics — in effect, a template by another name.

These issues tend to feed the theoretical debate as to whether there is or is not a post-syntactic ‘morphological component’ or whether syntactic mechanisms alone are sufficient to get clitics into the right order. Some syntactic movement accounts also rely on post-syntactic ‘morphological adjustment’ mechanisms as well, leaving morphology with the familiar role as the ‘scrap heap’ of syntactic theory, i.e. the place where syntacticians send inconvenient facts that do not fit into elegant movement-based analyses. Depending on one’s views on the place of morphology in the grammar, and what sorts of elements and principles it might include, this approach may be more or less attractive, but it leaves the whole problem of clitic ordering intact.

There is of course another recent approach, which undertakes to bridge the divide between syntax and morphology by treating the ordering of elements in both modules by means of constraints on alignment. The Optimality Theoretic approach to clitic ordering, pioneered by Anderson (1996) and later taken up by Grimshaw (1997, 2001), uses constraints from the ALIGN family to place clitics (like all other morphemes) in a given order with respect to one another. The problem with these accounts is that the constraints which are posited are completely unmotivated, other than by the attested ordering facts. So in a sense the OT accounts suffer from the worst of both the syntactic-movement based analyses and the base-generation morphology analyses: clitic ordering is portrayed as the result of otherwise unmotivated ad hoc constraints which remain both ungeneralizable and non-predictive. The OT approach to clitic ordering amounts to a notational variant of the template formalism, with the same drawbacks but perhaps the (unintended) virtue of making the over-predictive power of the ‘constraints’ it necessarily posits even more obvious.

It should be noted that virtually all contributions to the debate over the last three and a half decades accept unquestioningly the generalization in 2 (i.e. the fixed order of clitic sequences) as one of the ‘facts’ that must be accounted for in any analysis. Perlmutter notes (1971: 50-51, note 28) that there is some variation with respect to whether speakers from different Spanish dialects accept sequences of just two clitics (like those in 3a.) or longer sequences of three or four clitics (Bastida 1976). This variation is

however taken as further evidence for the strict ordering of clitics i.e. whatever their length, clitic sequences have to follow the template in 1. Despite this widely accepted ‘fact’, many speakers of Spanish in different countries hear (and some regularly produce) recurrent ‘incorrect’ forms like:

4. Me se secó todo. (cf. Standard Se me secó todo)
Te se apaga el fuego. (cf. Standard Se te apaga el fuego)

Such sequences are of course violations of 1 and something which the Spanish prescriptive norm condemns in no uncertain terms:

5. Esta construcción es estimada en todas partes como solecismo plebeyo.
(Real Academia Española 1973:427)

This proscription is of course strong evidence that lots of people actually use these nonstandard orders, as is the mnemonic phrase in 6., widely reported to have been used in Spanish schools to help reinforce the ‘correct’ order, and the corrections in 7, carried out automatically by a well-know word-processing program:

6. Como la semana antes del mes, *se* antes de *me*.

7. *me se* > *se me*
te se > *se te* (MS Word TM Spanish grammar-checker)

Note that none of the other ungrammatical sequences in 3b get this kind of treatment from the prescriptive authorities: while they (and many other mathematically possible but unattested sequences) are equally bad from the point of view of the template in 1, they are not explicitly banned by prescriptive grammarians, nor corrected by word-processors nor actively drummed out of children in school. If some violations of the clitic template are the object of special normative pressure, this most likely means that these

particular nonstandard clitic combinations (unlike others) actually occur in natural speech with some frequency. And so Perlmutter's overly-sweeping statement in 2 seems less like a 'fact' and more an idealization, which could be reframed more realistically as something like:

8. Clitics are *for the most part* strictly ordered, so we might as well treat them as if they were all strictly ordered. (= restatement of 2, my italics).

The idealization in 8 is of course quite understandable, and may even seem like a necessary methodological simplification. It should be noted that we are *not* talking about variation in clitic ordering everywhere in all possible sequences (since there are no normative injunctions against sequences in 3b we can safely assume that such variants do not occur with sufficient frequency to worry about), but rather a very localized phenomenon which affects just the first and second person singular clitics *me* and *te* with respect to the third person reflexive *se*: interestingly, speakers who have the nonstandard orders as in 4 rarely if ever accept the plural equivalents (**nos se*, **os se*) and often vary with great freedom between the standard and nonstandard orders (*me se* / *se me*, *te se* / *se te*).¹ So the generalization in 8 has the status of a 'near-universal' i.e. something that is almost completely true, which might make it seem quite reasonable to just ignore the small number of clitic sequences which might not always be strictly ordered. Note that I am not claiming that anyone deliberately ignored the facts about variable sequences, even though at least some people involved must have known about some of the possible variation in clitic orders. Rather I am suggesting that a collective, tacit agreement along the lines of 8 and 1 must have seemed like a fair idealization of the facts upon which to base analyses and debate, setting aside the small amount of attested variation (as perhaps 'marginal') at least until we figured out what was going on with the rest of the (strictly ordered) clitics.

But thirty-seven years (and counting) after Perlmutter introduced it, just where has this idealization gotten us? We still don't have a principled unified account of clitic

¹ Inés Fernández-Ordóñez (p.c.) cites a few isolated cases of *nos se* in the *COSER* database, and one unconfirmed report of *os se*, but these challenging and rare data fall outside the scope of this paper.

sequencing in Spanish, so maybe it is time to revisit that comfortable idealization and look at what people actually say instead of what they are supposed to say. Idealizations are all very well, even necessary at times, but only as long as we keep in mind that they are not the same as facts, and that any account based on an idealization of speakers' grammar leaves out an element of variability which may in fact turn out to be at least as interesting as the idealized facts.

Nonstandard sequences like those in 4 are documented in dialect monographs on specific local speech varieties, but such monographs do not give us a general view of how widespread such structures are, for example. Fortunately, we have another source of data on nonstandard Spanish varieties, gathered some 40 years before Perlmutter wrote on clitic orders. Beginning in 1931, teams of dialectology fieldworkers criss-crossed the Iberian Peninsula asking rural speakers (mostly sedentary, older, male peasants) to repeat an extensive questionnaire in their local speech variety (see Navarro Tomas 1975, Heap 2002 and www.alpi.ca for details on the *Atlas lingüístico de la Península Ibérica*). Among the morphosyntactic questions in this survey were:

9. Se me cayó del bolsillo (I, 348)

¿Se te calmó el dolor? (I, 349)

Cuestionario I *Atlas lingüístico de la Península Ibérica* (Madrid 1930: 24).

Of course, from a modern perspective such a survey seems methodologically shocking: there are no grammaticality judgements, no 7-point Likert scale, no distracter sentences and no randomization of questions (there was also no such thing as 'informed consent', no waivers to sign and no Research Ethics Board or Institutional Review Board to satisfy).

These methods were of course the standard state-of-the-art in traditional dialectology or (as it was coming to be called) linguistic geography. What the dialect informants were asked to perform was in essence a bidialectal 'translation task': for each stimulus provided by the fieldworkers in something like Standard Spanish (for the regions that concern us here) they would give an equivalent in whatever vernacular

variety was spoken locally. One might reasonably expect that under such conditions, translation ‘priming effects’ would lead to the subjects producing a direct syntactic calque of the Standard Spanish stimulus, with perhaps some local pronunciation features and/or lexical items. But in fact what happened was quite the opposite: here, as elsewhere in traditional dialect studies of this sort, one finds many responses (for some questions, the majority) which differ significantly from the standard stimulus in systematic and therefore interesting ways. While some speakers might likely provide answers which copy the standard stimulus pattern, few if any are likely to deliberately answer with ‘invented’ forms which deviated from the standard except for those which they in fact actually use, a pattern which leads me to posit the working rule in 13:

10. Methodological postulate for geolinguistic data elicitation:

If there is any distortion or experimental error in a geolinguistic or dialectological survey, it will normally be in the direction of more standard forms: therefore, the nonstandard (or vernacular) forms which we find in the data probably reflect in reality a higher (not lower) rate of occurrence of the phenomenon in question.

The *ALPI* clitic ordering data from the 1930s, left untouched for almost 70 years (but now available at www.alpi.ca), provides us with an overview of the responses for nonstandard speakers who were born mostly in the second half of the 19th century. The most striking thing about the data in Map 1 is how overwhelmingly frequent the nonstandard orders were: of 234 points considered here (roughly speaking, the ‘Spanish’ survey area, excluding Galician-Portuguese and Catalan-Valencian areas as well as those points where the responses included only one proclitic or which had enclitics) only a tiny minority (3%) have standard orders (*se me*, *se te*) for both of the survey questions in 12. Not surprisingly, most of those who provide standard responses most often have at least one sociodemographic feature which distinguishes them from typical dialect survey subjects (i.e. Nonmobile Older Rural Males or NORMs): the standard orders usually correspond to points where for some reason the surveys were (exceptionally) conducted with a woman, a person with more than basic education and/or someone with an

occupation other than manual labourer. On the other hand, the overwhelming majority of the informants (75%) have both nonstandard orders (*me se*, *te se*) and correspond mostly to typical dialect subjects. These two apparently invariant grammars (completely nonstandard or completely standard) do not necessarily pose a major challenge, at least initially, for the description of clitic orderings. For the nonstandard orders we could postulate a template similar to the one in 1 but with *se* placed after the first and second persons:

11. Modified output condition on clitic pronouns:

| | | | |
|----|---|-----------|-----|
| II | I | <i>se</i> | III |
|----|---|-----------|-----|

The template in 11 would however incorrectly predict that first and second person plural clitics also precede the third person reflexive (**nos se*, **os se*), something which is not in fact attested in these varieties (García 1960). So the template would have to be further complicated in order to distinguish between both grammatical persons and number:

12. Re-modified output condition on clitic pronouns:

| | | | | | |
|------|------|-----------|------|-----|-----|
| IIsg | I sg | <i>se</i> | IIpl | Ipl | III |
|------|------|-----------|------|-----|-----|

Alternatively, one could equally posit the equivalent inversion in the orders of whichever syntactic movements to whatever functional heads you happen to use for *se me*, *se te*, or in the relative ranking of the respective ALIGN constraints in an OT tableau. Given that the resulting analysis has to distinguish both grammatical person and number, it is bound to be rather messy and ad hoc, no matter which framework is used.

Now apart from these two categorical grammars, there are two other types of grammars shown in Map 1 which are not categorical but rather mixed or ‘hybrid’ in interesting ways. There are a small number (barely 2%) which show variation between both standard and nonstandard variants: *me se* / *se me*, *te se* / *se te*: the beauty (and messiness) of working with original survey materials is that the fieldworkers noted when informants provided more than one response. A rather larger group (some 20%) shows a ‘mixed’ response: *se me* for line 348 and *te se* for line 349. Interestingly, the reverse combination is never attested: the nonstandard sequence for the first person singular (*me*

se) does not seem to co-occur with the standard sequence for second person singular (*se te*). Which leads to the following implicational generalization:

13. If a nonstandard grammar of Spanish allows a clitic to precede *se*, it can be either *te* alone, or one of *te* or *me*, but not just *me* alone.

From the data in Map 1 we can see a marked asymmetry in these two nonstandard sequences: *te se* occurs at more points than *me se*.

How exactly to represent such an asymmetry is not among the goals of this paper;² my purpose here is not to propose any particular analysis but simply to point out the dangers of ignoring such nonstandard data. It should be noted that that in any of the competing frameworks typically used to analyse clitic sequences (syntactic movement, template morphology or OT constraints) the mechanisms needed to account for such data would have to be highly ad hoc, to say the least. This fact alone should lead us to consider seriously whether we have in fact been pursuing the right sorts of analyses for clitic ordering phenomena in general.

It would however be prudent to consider the possibility that the ‘problem’ lies with the *ALPI* data collected in the 1930s rather than with available accounts for clitic linearization. The facts in Map 1 could of course turn out to be nothing more than a blip due to methodological noise of one kind or another. One source available for comparison is another series of linguistic atlas surveys conducted in Spain between the 1950s and the 1980s, in Andalusia (*ALEA*), Aragon, Navarre and Rioja (*ALEANR*), Cantabria (*ALECan*) and the region of Castilla y Leon (*ALCyL*). This series of atlases used a questionnaire which included the elicitation of the same two clitic sequences as in 9. In Figure 1 the data extracted from these atlases are given in aggregate form, showing in each case the percentage of nonstandard forms with the first person singular clitic (*me se*) and the percentage of with the second person singular (*te se*).

² Elsewhere (Heap 2005) I discuss how such an asymmetry might be modeled using differences in structural markedness, i.e. through a morphological feature geometry,

| | <i>me se</i> | <i>se me</i> | both | % nonstandard | <i>te se</i> | <i>se te</i> | both | % nonstandard |
|---------------|--------------|--------------|------|---------------|--------------|--------------|------|---------------|
| <i>ALEA</i> | 203 | 23 | 4 | 90,0 | 207 | 19 | 4 | 91,7 |
| <i>ALEANR</i> | 109 | 43 | 23 | 75,4 | 118 | 34 | 23 | 80,6 |
| <i>ALECan</i> | 9 | 40 | 6 | 27,3 | 9 | 40 | 6 | 27,3 |
| <i>ALCyL</i> | 76 | 105 | 19 | 47,5 | 86 | 95 | 19 | 52,5 |

Figure 1. Standard vs. Nonstandard clitic sequences from regional linguistic atlas data

Although in one case (the *ALECan* data) the same proportion (27.3%) is given for both persons, this is because these data are presented on the same map without distinguishing the two different persons. Apart from that, the same generalization comes out: there is always a greater proportion of *te se* than *me se*. So the implicational generalization in 13 seems to be born out here as well: even taking into account the *ALECan* case, we can still say that the proportion of nonstandard sequences for *me se* is never higher than for *te se*.

Of course, all these atlas surveys used a similar methodology, with a closed questionnaire and ‘translation task’ responses, and so it would be reasonable to wonder if these data were somehow an artifact of the elicitation techniques. More recently, using: semi-directed interview techniques, the *COSER* (Corpus Oral Sonoro del Español Rural, Fernández-Ordóñez 2005) survey has been collecting more spontaneous data from contemporary nonstandard varieties. While such an interview situation can never be fully naturalistic, there is no set list of structures to be elicited and as such there is much less likelihood of experimenter-induced skewing of morphosyntactic structures. The first thing to be noted is that the target structure (*se* in combination with *te* and / or *me*) is in itself not all that frequent in this corpus. It typically arises from the combination of a reflexive verb (*olvidarse*, *irse*, *perderse*) with an ethical dative marking the affectedness

of the speaker (*me*) or their interlocutor (*te*). Since ethical datives are always grammatically optional (if omitted the sentence remains grammatical and loses only the nuance of affectedness), it is quite possible to go through many hours of interview data without a speaker ever using this structure, which implies a feeling of involvement with the utterance. Not surprisingly then, when these combinations arise in interviews they are more common in the first person (*se* with *me*) than in the second (*se* with *te*) since a speaker is less likely to attribute affectedness to an interviewer with whom they are barely acquainted.

| Provincias: | <i>se me</i> | <i>me se</i> | % | <i>se te</i> | <i>te se</i> | % |
|---|--------------|--------------|--------------|--------------|--------------|--------------|
| Cantabria, País Vasco, Navarra | 28/29 | 1/29 | 3,4% | 17/18 | 1/18 | 5,6% |
| Burgos, La Rioja, León, Palencia, Soria, Valladolid, Zamora | 45/52 | 7/52 | 13,5% | 11/18 | 7/18 | 38,9% |
| Ávila, Salamanca, Segovia, Madrid, Teruel, Toledo. | 81/90 | 9/90 | 10,0% | 40/51 | 11/51 | 21,6% |
| Badajoz, Cáceres, Ciudad Real. Jaén | 6/9 | 3/9 | 33.3% | 3/7 | 4/7 | 57,1% |

Figure 2. Standard vs. nonstandard clitic sequences from the *COSER* database

Given the overall infrequency of the target structure, Figure 2 groups together interviews from a number of different provinces in order to get enough data to see overall trends. What is striking here is that wherever the nonstandard sequences are used, *te se* is proportionally more frequent (as a percentage of all the tokens with the second person) than *me se* (among all first person tokens), even though the first person occurs more frequently in absolute terms. That is to say, exactly the same implicational asymmetry between the two grammatical persons which we see in the atlas questionnaire data occurs in these more spontaneous interview data, even as the nonstandard sequences become less frequent overall.

The diachronic trend over the last century is also very clear: once the variants used by the overwhelming majority of less-educated speakers born in the second part of the 19th century (*ALPI* data), the nonstandard sequences (*me se*, *te se*) become less common

throughout the 20th century (as shown by the other linguistic atlas survey data in Figure 2) and finally towards the turn of the 21st century they are relegated to minority status in virtually all regions, and this in the speech of elderly rural subjects chosen for their ties to traditional speech communities. Clearly the normative proscription and stigmatization of the vernacular variants has been successful in all but stamping out the nonstandard sequences. While it is tempting to see this rapid change as a recent and massive change in progress, diachronic corpus data show very few occurrences of the nonstandard variants. Apart from usage in *costumbrista* novels of the 18th and 19th century as a marker of ‘local colour’, typically in dialogue attributed to uneducated peasants, there are very few nonstandard sequences from Spanish texts in the 12th-16th centuries: the prescriptive norm appears to have been highly successful in keeping these vernacular forms out of the written language. Strikingly, however, where the nonstandard sequences do occur, although the total numbers may be vanishingly small, the same asymmetries shown above apply: *te se* can occur without *me se* being present, and if they both occur then *te se* is proportionally more frequent than *me se* (Heap 2006).

A renowned historian of the Spanish language notes the same asymmetry in a passing comment, while confirming the normative stigmatization of both nonstandard sequences: “Totalmente inculta es la anteposición de *me* y *te* a *se* («*me se cayó*», «*te se olvida*»), aunque el *te se* cuenta con cierta indulgencia en algunas regiones.” (Lapesa 1980: 472). While he tells us nothing about the data which form the basis for his observations, Lapesa seems to have grasped (intuitively?) a subtle difference between the two nonstandard variants which corresponds to the same asymmetry that comes out of the dialect data presented above.

What are the implications of all this variation for morphosyntactic theory? We could of course represent the categorically nonstandard variants (i.e. those grammars that have only and always *me se* and *te se*) using a modified output template as in 12 or its equivalent in syntactic movements or OT constraints: the resulting ad hoc analyses would reveal and predict as much (or rather, as little) about clitic sequences as the mechanisms which have been proposed for the standard sequences. Crucially, there seem to be no relevant correlations between these differences in clitic orders and other aspects of

syntactic movement, other ALIGN constraints or other elements that might be affected by a template. Furthermore, the mixed grammars which allow both standard and nonstandard variants, and in particular the person asymmetries between *me* and *te* when combined with *se*, should lead us to reconsider whether these various competing approaches to clitic linearization are really up to the task. Any descriptively adequate account, whether based on syntactic movement, ALIGN constraints or output templates, must necessarily make crucial reference to the morphological makeup of the clitics involved, i.e. their person and number features. While this is of course possible (ALIGN constraints and syntactic heads can, if required, make reference to person and number features, just as templates can), such an increase in formal power seems theoretically undesirable unless accompanied by some more restrictive theory which directly addresses the internal morphological structure of the clitics themselves in some more enlightening fashion. Templates like those in 1 and 12 refer to morphological features, of course, but without any principled account of why the featural content of some particular clitic may lead to it being preposed or postposed (or both) with respect to another one. Whatever solutions are eventually proposed that may account for clitic ordering facts in a more principled way, it seems more than probable that they will have to involve a theory of pronoun-internal morphological structure which can structure possible clitic sequences.

Crucial to this enterprise is the availability of good morphosyntactic data for a range of vernacular varieties, such as is provided by linguistic atlases. Other Romance language areas such as French and Italian may have a longer and better-documented history of research using linguistic atlases data, but even there the necessary data may not be easy to find:

A parte la loro scarsità oggettiva, va sottolineato che i dati sintattici compaiono negli atlanti [linguistici] in modo casuale; un aspetto non marginale del problema sta quindi nel fatto che anche i dati presenti sono in genere difficilmente reperibili, in quanto si trovano nelle ‘pieghe’ di dati ricercati per scopi diversi. (Benincà 1992: 29).

One advantage of the *ALPI* questionnaire (see www.alpi.ca/questionnaire.php) is that it was clearly designed in part with morphosyntactic variables in mind: empirical

phenomena which are investigated systematically include pseudo-reflexive or impersonal *se*, agreement of ‘invariant’ *haber*, variation in direct and indirect object pronouns, among many others (Heap 2002). These and other examples of data from geolinguistic sources like the *ALPI* surveys which pose interesting challenges for formal analyses of phenomena such as ‘agreeing’ adverbs (Pato & Heap 2005) or the distribution of perfective vs. preterit verbal forms (Carter 2003; Heap & Pato 2006). Further evidence of the usefulness of dialect variation data for morphosyntactic analyses can be seen in the work of Fernández-Ordóñez (1994, 1999, 2006, 2007) and Enrique-Arias (2006, 2007). My purpose here however is not to explore all of these empirical issues but rather to point out a more general methodological issue: morpho-syntactic theory should take more seriously vernacular data from linguistic geography, not only because (or despite) of the descriptive challenges it may provide, but more crucially because sometimes vernacular data may in fact provide important insights into what a more predictively adequate solution might entail. Such alternate directions may not be apparent if analyses are based exclusively on standard facts: i.e. if seemingly messy facts are not ‘idealized out’ of the dataset from the outset. Conversely, research in linguistic geography can achieve more focused results with greater impact outside of this particular subfield if issues in morphosyntactic theory are used as a lens through which facts from vernacular varieties are viewed. Both sides of this divide need (and ought to listen to) each other.

References

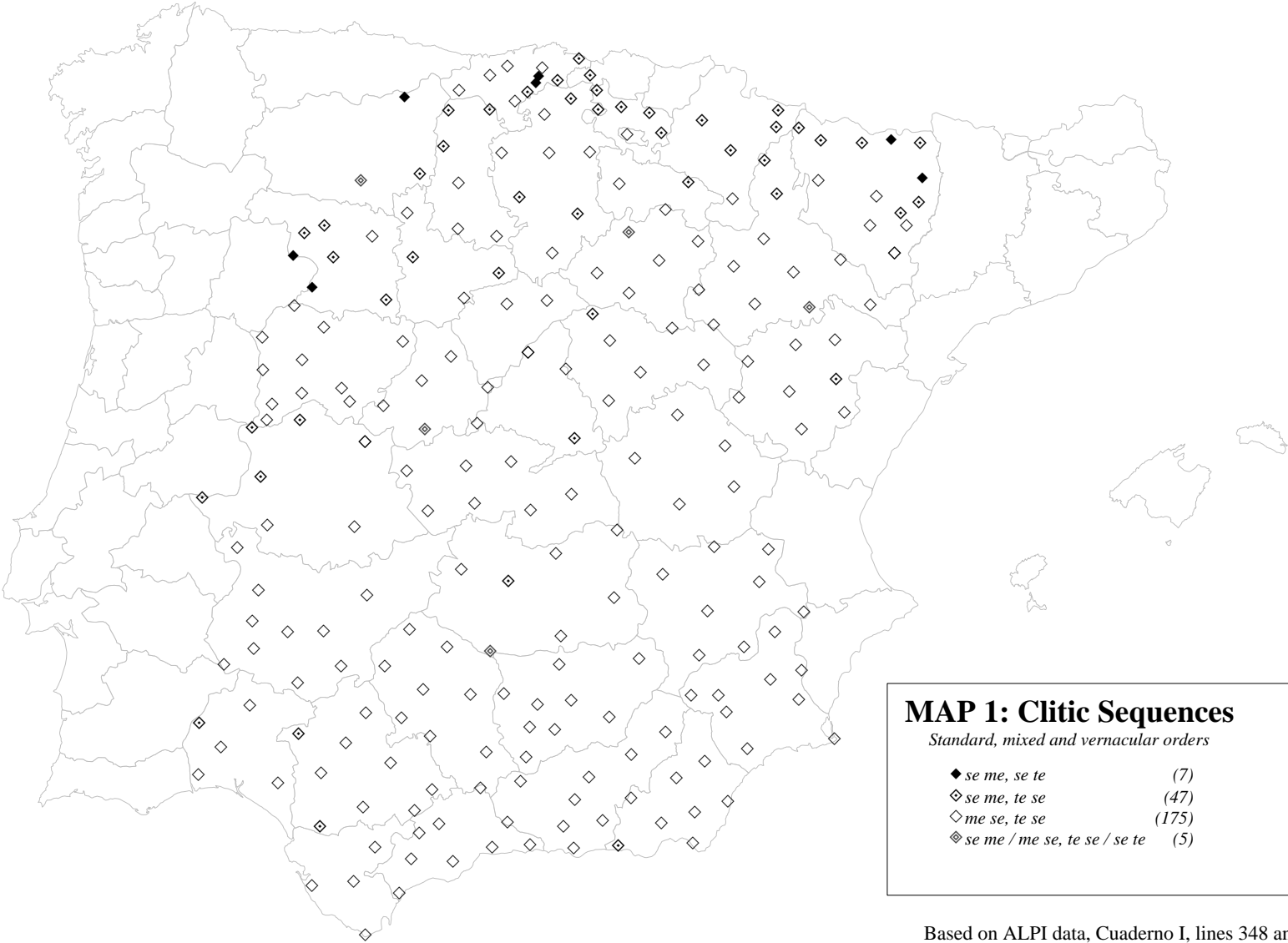
- ALPI: Atlas lingüístico de la Península Ibérica*. Madrid: CSIC, 1962. Materiales en línea: www.alpi.ca University of Western Ontario, London, Ontario, Canadá. David Heap, director, 2001-.
- ALVAR, Manuel (1995) *Atlas lingüístico y etnográfico de Cantabria*, Madrid: Arco Libros, 2 tomos.
- ALVAR, Manuel. (ed.) (1999) *Atlas Lingüístico de Castilla y León*, Valladolid: Junta de Castilla y León/ Fundación “Duques de Soria”.

- ALVAR, Manuel, Antonio LLORENTE & Tomás BUESA (1979-1983) *Atlas lingüístico y etnográfico de Aragón, Navarra y Rioja*, Madrid: Departamento de Geografía Lingüística, Institución Fernando el Católico y el Consejo Superior de Investigaciones Científicas.
- ALVAR, Manuel, Antonio LLORENTE & Gregorio SALVADOR (1961-1973) *Atlas lingüístico y etnográfico de Andalucía*, Granada: Universidad de Granada.
- ANDERSON, Stephen R. (1996) "How to put your Clitics in their Place, or Why the Best Account of Second-Position Phenomena May be a Nearly Optimal One", *The Linguistic Review*, 13, 165-191.
- BASTIDA, Salvador (1976) "Restricciones de orden en las secuencias de clíticos del castellano: dos requisitos". In V. Sánchez de Zavala (ed.): *Estudios de gramática generativa*, Barcelona: Labor, 59-99.
- BENINCÀ, Paola (1992) "Note introduttive a un atlante dialettale sintattico", in M. Cortelazzo (ed.), *Dialectologiae varia linguistica*, Padua: Unipress.
- BONET, Eulàlia (1991) *Morphology after syntax: Pronominal clitics in Romance*, Ph.D. dissertation, MIT.
- BONET, Eulàlia (1994) "The Person-Case Constraint: A Morphological Approach", *MITWPL 22 The Morphology-Syntax Connection*, 33-52.
- BONET, Eulàlia (1995) "The Where and How of Clitic Ordering", *Revue québécoise de linguistique*, 24.1, 61-81.
- CARTER, Diana (2003) *Variation in the present perfect, based on the Atlas Lingüístico de la Península Ibérica (ALPI)*. MA thesis: University of Western Ontario.
- DINNSSEN, Daniel (1972) "Additional constraints on clitic order in Spanish", in B. Saciuk & J. Casagrande (eds.), *Generative Studies in Romance Languages*, Rowley, MA: Newbury House, 175-183.
- EMONDS, Joseph (1975) "A Transformational Analysis of French Clitics without Positive Output Constraints", *Linguistic Analysis* 1, 3-24.
- ENRIQUE-ARIAS, Andrés (2006) *Realizaciones del clítico os en el ALPI*. Madrid: Sociedad española de lingüística.

- ENRIQUE-ARIAS, Andrés (2007) *Motives for change in the paradigm of Spanish object clitics*. International Conference on Historical Linguistics. UQÀM, Montréal. August 2007.
- FERNÁNDEZ-ORDÓÑEZ, Inés (1994) “Isoglosas internas del castellano. El sistema referencial del pronombre átono de la tercera persona”, *Revista de filología española*, 74, 71-125.
- FERNÁNDEZ-ORDÓÑEZ, Inés (1999) “Leísmo, laísmo y loísmo”, in I. Bosque & V. Demonte (dirs.), *Gramática Descriptiva de la Lengua Española*, 3 volumes, Madrid: Espasa-Calpe, 1317-1397.
- FERNÁNDEZ-ORDÓÑEZ, Inés (2005) *COSER: Corpus oral y sonoro del español rural*. <http://pidweb.ii.uam.es/coser/>
- FERNÁNDEZ-ORDÓÑEZ, Inés (2006) “Del Cantábrico a Toledo: El “neutro de materia” hispánico en un contexto románico y tipológico. (primera parte)”, *Revista de Historia de la Lengua Española*, 1, 67-118.
- FERNÁNDEZ-ORDÓÑEZ, Inés (2007) “Del Cantábrico a Toledo: El “neutro de materia” hispánico en un contexto románico y tipológico (segunda parte)”, *Revista de Historia de la Lengua Española*, 2, forthcoming.
- GARCÍA MARTÍNEZ, G. 1986 (1960). *El habla de Cartagena: palabras y cosas (Notas para el estudio del castellano vulgar actual y de la propagación del aragonés y del catalán por el Sur)*, Murcia: Universidad de Murcia.
- GRIMSHAW, Jane (1997) “The Best Clitic: Constraint Conflict in Morphosyntax”, in L. Haegeman (ed.) *Elements of Grammar*. Dordrecht: Kluwer, 169-196.
- GRIMSHAW, Jane (2001) “Optimal Clitic Position and the Lexicon in Romance Clitic Systems”, in G. Legendre, J. Grimshaw & S. Vikner (eds.), *OT Syntax*, Cambridge, USA: MIT Press.
- HARRIS, James (1994) “The syntax-phonology mapping in Catalan and Spanish clitics”, in A. Carnie & H. Harley (eds.), *MITWPL*, 21:321-353.
- HARRIS, James (1996) “The morphology of Spanish clitics”, in H. Campos & P. Kempchinsky (eds), *Evolution and Revolution in Linguistic Theory: Essays in Honor of Carlos Otero*, Washington D.C.: Georgetown University Press.

- HEAP, David (2002) “Segunda noticia histórica del ALPI a los 40 años de la publicación de su primer tomo”, *Revista de filología española*, LXXXII, 5-19.
- HEAP, David (2005) “Constraining Optimality: Clitic Sequences and Feature Geometry”, in L. Heggie & F. Ordóñez (eds), *Perspectives on Clitic and Agreement Affix Combinations* (Linguistik Aktuell series), Amsterdam: John Benjamins, 81-102.
- HEAP, David (2006) *Secuencias ‘invertidas’ de pronombres clíticos: un cambio (¿?) en tiempo real*, Actas del Congreso Internacional de la Historia de la Lengua Española. Madrid, Arcos Libros. Vol. 1, 785-799.
- HEAP, David & Yves ROBERGE (2001) “Cliticisation et théorie syntaxique, 1971-2001”, *Revue québécoise de la linguistique* 2002, 30:1, 63-90.
- HEAP, David & Enrique PATO (2006) *La organización dialectal del castellano: la distribución de ‘he cantado’ en la Península Ibérica*, México: VII Congreso Internacional de Historia de la Lengua Española: Mérida.
- HERSCHENSOHN, Julia (1980) “On clitic Placement in French”, *Linguistic Analysis* 6-2, 187-219.
- LAENZLINGER, Christopher (1993) “A syntactic view of Romance pronominal sequences”, *Probus*, 5:3, 241-270.
- LAPESA, Rafael (1980) *Historia de la lengua española*. Madrid: Gredos.
- NAVARRO Tomás, Tomás (1975) “Noticia Histórica del ALPI”, in *Capítulos de Geografía Lingüística de la Península Ibérica*, Bogotá: Instituto Caro y Cuervo, 9-20.
- PATO, Enrique & David HEAP (2005) “The lexical-syntactic process of recategorization: the case of Spanish medio”, in Claire Gurski, ed. *Proceedings of the Canadian Linguistic Association Annual Conference*, London: The University of Western Ontario. <http://ling.uwo.ca/publications/CLA-ACL/CLA-ACL2005.htm>
- PERLMUTTER, David (1971) *Deep and Surface Structure Constraints in Syntax*, New York: Holt & Rhinehart.
- PEARCE, Elizabeth (1991) “On comparing French and Italian: The switch from *ilium mihi* to *mihi ilium*”, in *New Analyses in Romance Linguistics*, Selected papers from the XVIII Linguistic Symposium on Romance Languages, Amsterdam: Benjamins.

- REAL ACADEMIA ESPAÑOLA (1973) *Esbozo de una nueva gramática de la lengua española*, Madrid: Espasa Calpe.
- SZABO, Robert (1974) “Constraints on clitic insertion in Spanish”, in R. Campbell et al. (eds.), *Linguistic Studies in Romance Languages*, Washington D.C.: Georgetown University Press, 124-138.
- URIAGEREKA, Juan (1995) “Aspects of the Syntax of Clitic Placement in Western Romance”, *Linguistic Inquiry*, 26, 79-123.
- WANNER, Dieter (1977) “On the Order of Clitics in Italian”, *Lingua*, 42, 101-128.
- WANNER, Dieter (1994) “El orden de los clíticos agrupados en castellano”, *Thesaurus, Boletín del Instituto Caro y Cuervo*, 48, 1-57.



Based on ALPI data, Cuaderno I, lines 348 and 349.
© 2008 www.alpi.ca