STRATEGY COMPETITION IN THE EVOLUTION OF PRONOUNS: A CASE-STUDY OF SPANISH *LEÍSMO, LAÍSMO* AND *LOÍSMO*

REMI VAN TRIJP
Sony Computer Science Laboratory Paris
6 Rue Amyot
Paris, 75005, France
remi@csl.sony.fr

Pronouns form a particularly interesting part-of-speech for evolutionary linguistics because their development is often lagging behind with respect to other changes in their language. Many hypotheses on pronoun evolution exist – both for explaining their initial resilience to change as well as for why they eventually cave in to evolutionary pressures – but so far, no one has proposed a formal model yet that operationalizes these explanations in a unified theory. This paper therefore presents a computational model of pronoun evolution in a multi-agent population; and argues that pronoun evolution can best be understood as an interplay between the level of language strategies, which are the procedures for learning, expanding and aligning particular features of language, and the level of the specific language systems that instantiate these strategies in terms of concrete words, morphemes and grammatical structures. This claim is supported by a case study on Spanish pronouns, which are currently undergoing an evolution from a case- to referential-based system, the latter of which there exist multiple variations (which are called *leísmo*, *laísmo* and *loísmo* depending on the type of change).

1. Introduction

Pronouns are commonly defined as substitutes for noun phrases that refer to persons or things in the discourse context. Interestingly enough, there is often a discrepancy between the morphosyntactic qualities of pronouns and the nominal system in the rest of the language (Comrie, 2005). In English, for example, full noun phrases are not marked for case, whereas pronouns distinguish between nominative and non-nominative (e.g. *he* vs. *him*). In most cases, the discrepancy means that pronouns still contain traces of an earlier grammatical stage of a language. This makes them interesting objects of study for evolutionary linguistics: not only because they hint at what the language once looked like, but also because they *"cannot indefinitely uphold a grammatical category/property-based distinction alone [...]"* (Howe, 1996, p. 63). In other words, pronoun evolution is strongly motivated by other, well-known developments in their language.

Many hypotheses have been proposed to explain why pronouns are initially resilient to change and why they eventually yield to evolutionary pressures. However, no one has proposed a formal model yet that operationalizes these proposals

and validates them in a computational model. This paper therefore presents a concrete multi-agent model that investigates the evolution of pronoun systems through a case study of Spanish, in which the pronouns are currently shifting from a case- to referential-based system (Valenzuela et al., to appear). The model studies pronoun evolution at two different levels (Bleys & Steels, 2009; Steels, 2009):

1. The level of *language strategies*: Language strategies are sets of procedures for acquiring, expanding and aligning features of language. Language strategies allow speakers to become and remain proficient in their language.

2. The level of *language systems*: Language systems are the concrete choices that instantiate particular language strategies in terms of an ontology, a lexicon and grammatical structures. Language systems allow language users to produce and parse conventional linguistic utterances.

As I will illustrate in the remainder of this paper, there is an interplay between both levels that strongly influence the ongoing evolution in Spanish.

2. A Case Study of Spanish Pronoun Evolution

The (Standard) Spanish personal pronouns still show traces of a previous case declension. This is most apparent in the third person, which differentiates among nominative, accusative and dative. The other pronouns, however, only distinguish between nominative and non-nominative. The third person pronoun also features a gender distinction between masculine and feminine in the nominative and accusative cases, but not in the dative case. This pronoun system (see Table 1) is known by scholars as the 'etymological system' (Valenzuela et al., to appear).

Table 1. Singular pronouns in Standard Spanish.

	Nominative	Accusative	Dative
1st Person	*yo*	*me*	*me*
2nd Person	*tú*	*te*	*te*
3d Person Masc.	*él*	*lo*	*le*
3d Person Fem.	*ella*	*la*	*le*

The etymological system is gradually changing into a 'referential system' in which the accusative and dative cases are collapsed. This doesn't mean that the system is simply impoverished: pronouns start to differentiate gender, number and noun class. Moreover, the pronouns have specialized into agreement markers as they can be used in a sentence along with the noun phrases they refer to:

(1) *Le* *di* *un regalo* *a mi madre.*
him.3SG.DAT. give.1PRES a gift to my mother
'I gave my mother a present.' (*lit.* 'to her I have a present to my mother')

'Referential system' is in fact an umbrella term for covering multiple variations of the system that are not uniformly propagated in the regions of Spain, and the referential system is still in competition with the etymological one (Fernández-Ordóñez, 1999). Depending on the particular flavour of the referential system, the variations are called *leísmo*, *laísmo* and loísmo.

Leísmo denotes the use of the pronoun *le* (etymologically a dative pronoun) instead of *lo*. The most frequent occurrences are the use of *le* as a singular, masculine and personal pronoun (ex. (2)). The *Laísmo* variation concerns the use of the pronoun *la* (etymologically an accusative pronoun) instead of *le* with a feminine referent (ex. (3)). Finally, Loísmo is the use of *lo* (etymologically an accusative pronoun) instead of *le* with masculine or neuter referents (ex. (4)).

(2) *Le* *vi* *(a* *Javier).*
her.ACC saw-1P.SG.PAST (Person-Marker Javier)
'I saw her (Javier).'

(3) *La* *dio* *un* *regalo* *(a* *Maria).*
her.DAT gave-3P.SG.PAST INDEF.ART.M present to Maria
'He gave a present to her (Maria).'

(4) *Lo* *dio* *un* *regalo* *(a* *Juan).*
him.DAT gave-3P.SG.PAST INDEF.ART.M present to John
'He gave a present to him (John).'

3. Two Levels of Language Evolution

This paper assumes that the evolution described in the previous section is not happening by accident but that it is *motivated* given the other developments in the Spanish language, which has lost case marking in its noun phrases (in favour of word order) and in which explicit gender-marking has become increasingly important (Valenzuela et al., to appear). Spanish pronoun evolution should therefore be studied at two levels: the level of the *language system* and the level of the *language strategies*. Both levels are studied within the setting of a *language game*.

3.1. *Language System*

Studies of grammaticalization have abundantly shown that language users most of the times recycle existing forms into new functions. This is also true for the Spanish pronouns. Future innovations and changes in a language are therefore highly dependent on the already existing items in the linguistic inventory.

The experiments therefore start with a population of agents that are equipped with the etymological pronoun system of Spanish, including an ontology, lexicon and grammar. The main challenge of this formalization is that speakers of Spanish are perfectly capable of recognizing and understanding when other speakers use

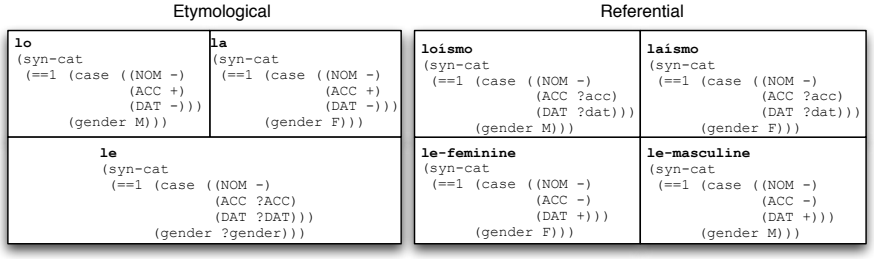

Figure 1. *Left:* A simplified representation of the syntactic functions of *lo*, *la* and *le* in the etymological system. *Right:* The possible syntactic functions of the same pronouns in the referential system.

a particular variation such as *laísmo* or *loísmo*, so the agents have to be capable of dealing with variation as well and be aware of alternative uses. This has been achieved in Fluid Construction Grammar (De Beule & Steels, 2005; Steels & De Beule, 2006, also see `www.fcg-net.org`). An overview of the current implementation can be found in Valenzuela et al. (to appear).

Ontology. The experiments make use of a small ontology that consists of several objects and events. The objects differentiate between 'persons' (e.g. [JAVIER]) and 'non-persons' (e.g. [BOOK]). This distinction is not relevant for the experiments in this paper, but will be used for later stages of the case study (also see section 4). There are in total 28 objects that balance the gender distinction between masculine-feminine. Next, there are eight different events, of which five transitive (e.g. [TOUCH]) and three ditransitive events (e.g. [GIVE]).

Lexicon and Grammar. Each item in the ontology has a corresponding lexical entry (e.g. [CAT] – "gato"). The grammar consists of two argument-structure constructions for transitive and ditransitive events. These constructions take care of the correct word order and grammatical role assignment in production. In parsing, they help the hearer to retrieve 'who did what to whom'.

The pronouns *lo, la* and *le* are implemented as morphological rules that map a specific syntactic function to a particular form. The left of Fig. 1 illustrates the case and gender specification of the etymological system. The case-features of *lo* and *la* specify that they always play the accusative role (marked with a '+') and never the nominative or dative roles (marked with a '–'). The two pronouns only differ in the feature 'gender', which has the value 'M' (masculine) for *lo*, and 'F' (feminine) for *la*. *Le*, on the other hand, states that it always plays the dative role. However, it leaves its gender underspecified, which is marked by the

variable '?gender' (indicated with a question mark). Indeed, the utterance *Le dio un regalo* will be ambiguous as to whether the speaker intended to say 'He gave **him** a present' or 'He gave **her** a present'.

3.2. *Language Strategies*

The language system will allow agents to produce and parse conventional utterances in Spanish that involve the etymological pronoun system. However, becoming and remaining proficient in a language also means that you know how to acquire innovations, how to expand the system if needed and how to adapt your linguistic inventory to changes in the speech community. These three components are operationalized in the form of *language strategies* (Steels, 2009).

Each language strategy is explicitly represented as a set of 'diagnostics' for detecting communicative problems, 'repair functions' for solving these problems and 'alignment functions' for adapting the linguistic inventory based on the communicative outcome of an interaction. Each language strategy also has a 'score' between 0 and 1 that reflects its strength or dominance in the language. The dominance of a strategy will decide on which communicative problems are considered more urgent by the language user, which repair functions have priority over the other, and which alignment functions have the biggest impact.

A Strategy for Case. Given the loss of case marking in most of the Spanish language, I will assume here that the strategy for case is not productive anymore for Spanish speakers, and that it can at best be used for acquiring case distinctions and upholding this distinction because of frequency effects. The necessary learning mechanisms for a case strategy are described by van Trijp (2008).

A Strategy for Gender It is a well-established evolutionary pattern in Spanish that explicit gender-marking has become more and more important. The experiments therefore assume that the Spanish gender strategy has a high 'score' and that its associated diagnostics, repairs and alignment functions have a priority over those of the case strategy. In total, the gender strategy has three diagnostics with three corresponding repair functions and one alignment function:

1. **A speaker** can *detect* whether the pronoun they used explicitly marked gender or not. If not, they can *repair* this problem by either recruiting another pronoun (typically *lo* or *la*) or by adapting *le* (see below). However, two conditions have to be met: (a) the score of the gender strategy must be higher than the score of the case strategy; and (b) the 'functional load' of the recruited pronoun must be covered by at least one other system in the language. The latter requirement means that there must be some formal way that hearers can notice in some way that an innovation took place (see below).

2. **A speaker** can *detect* variation in the language when they know multiple possibilities for expressing the same meaning. This may cause an explosion of search effort during processing. The speaker can decide to *repair* this problem by introducing additional gender constraints on one of the pronouns (e.g. specify that *le* should only cover masculine referents).

3. **A hearer** can *detect* the novel use of a pronoun (e.g. *lo* or *la* in dative position) if the 'functional load' of case assignment is also carried by another system in the language. In this case, the word order specifications of the argument-structure constructions will effectively inform the hearer that there is a mismatch between the case specification of the pronoun and what was expected by the construction. The hearer can use this information for *repairing* the problem.

4. **Both speaker and hearer** will increase the confidence score of a pronoun if it was used in a successful interaction and punish pronouns that try to cover the same gender by decreasing their scores. In case of failure, only the score of the pronoun that was used is decreased.

Agents can autonomously test whether an innovation is incompatible with the old use of a pronoun. If that is the case (typically when introducing new constraints), then a new morphological rule is introduced that co-exists with the old one. If not, then the repair occurs directly in the recruited pronoun. The gender strategy can (but does not have to) lead to the four different morphological rules that are shown in the right of Fig. 1. Here we see that *lo* and *la* may evolve into non-nominative masculine and feminine pronouns, whereas *le* can remain a dative pronoun if it specializes in either masculine or feminine referents.

4. Experimental results and discussion

The above-mentioned language system and language strategies for case and gender were implemented in a population of ten agents that engage in a series of 'language games' (i.e. routinized communicative interactions). More specifically, agents play description games in which the speaker has to describe an event to the hearer. These events are randomly created by a scene generator based on the ontology of the agents. The speaker has to use a pronoun for one of the participants that play a role in the action. The game is a success if the hearer agrees with the description, and a failure if he disagrees. The agents will exploit this setting when adapting their language. For each language game, two agents are randomly picked to act as either the speaker or the hearer.

In all of the simulations, the population of agents succeeded in modifying their pronoun system such that it incorporates a novel gender distinction for indirect objects. The two graphs in Fig. 2 show examples of the global evolution between *lo*, *la* and *le* on the level of the language system. The Y-axis shows the percentage

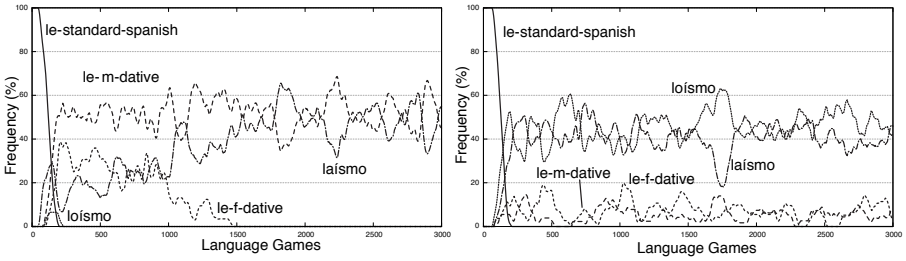

Figure 2. Two times competition between different uses of *le, la* and *lo*.

of how often a pronoun was used during the last ten interactions that involved a ditransitive utterance, whereas the X-axis indicates the number of interactions in the population.

The left graph shows a simulation where the agents went for the *laísmo* variation in which *la* emerges as a non-nominative feminine-marker. We see that the etymological use of *le* quickly disappears under influence of the new uses of *la* and also briefly of *lo*, but the form is nevertheless able to survive as a dative masculine pronoun. As the results show, whereas *le* used to take up 100% of the dative case, it now shares this function with *la* depending on the gender of the referent. The right graph shows an other simulation where *le* has almost completely disappeared and been replaced by *lo* and *la*, which makes that the accusative-dative distinction has completely collapsed. This case is actually the most frequent solution, as *le* is only able to survive the competition if it remains frequent enough.

It is striking that *leísmo* never occurs, whereas this is the most frequent variation in real life. This is due to the fact that there is already a gender differentiation in the accusative case, hence the gender strategy does not pressure the agents to make changes here. Indeed, most occurrences of *leísmo* in Spanish actually employ *le* as a pronoun for personal, animate and volitional referents (both masculine and feminine). This use of *le* probably follows a tendency in Spanish to differentiate person from non-person entities (Valenzuela et al., to appear). This hypothesis is supported by the form *a*, which emerged as a marker for person entities in direct object positions. The data therefore suggest that future experiments need to incorporate a third language strategy for marking the noun class or type.

5. Conclusions

This paper presented a multi-agent computational simulation in which a population of agents succeeded in evolving the etymological pronoun system of Spanish towards a more referential-oriented system, including the well-known *laísmo* and

loísmo variations of this system. The simulations showed that this shift was made possible without loss of communicative success because of evolution on two levels: that of the language system and that of language strategies. The results also suggest that gender can only explain two types of variation in Spanish and that at least a third strategy is needed in future experiments to explain some of the observed phenomena in Spanish.

Acknowledgements

This research was partly funded by the EU FP7 ALEAR Project and the Sony Computer Science Laboratory Paris. I would like to thank Javier Valenzuela, Mar Garachana Camarero and Joe Hilferty for their thoughtful comments. Many thanks also go to Luc Steels, director of Sony CSL Paris and the VUB AI-Lab in Brussels, and my colleagues from both labs for the helpful feedback on this work.

References

Bleys, J., & Steels, L. (2009). Linguistic selection of language strategies. a case study for colour. In *Proceedings of the 10th european conference on artificial life.* Berlin: Springer.

Comrie, B. (2005). Alignment of case marking. In M. Haspelmath, M. S. Dryer, D. Gil, & B. Comrie (Eds.), *The world atlas of language structures.* Oxford: Oxford University Press.

De Beule, J., & Steels, L. (2005). Hierarchy in Fluid Construction Grammar. In U. Furbach (Ed.), *Ki 2005: Advances in artificial intelligence. proceedings of the 28th german conference on ai* (Vol. 3698, pp. 1–15). Berlin: Springer.

Fernández-Ordóñez, I. (1999). Leísmo, laísmo, loísmo: Estado de la cuestión. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (Vol. I, pp. 1319–1390). Madrid: RAE – Espasa Calpe.

Howe, S. (1996). *The personal pronouns in the germanic languages. a study of personal pronoun morphology and change in the germanic languages from the first records to the present day.* Berlin: de Gruyter.

Steels, L. (2009). Can evolutionary linguistics become a science? *Journal of Evolutionary Linguistics*, *1*(1).

Steels, L., & De Beule, J. (2006). Unify and merge in Fluid Construction Grammar. In P. Vogt, Y. Sugita, E. Tuci, & C. Nehaniv (Eds.), *Symbol grounding and beyond.* (pp. 197–223). Berlin: Springer.

Valenzuela, J., van Trijp, R., Garachana, M., & Hilferty, J. (to appear). Variation in grammar and fluid construction grammar: A case study of the spanish clitic system.

van Trijp, R. (2008). The emergence of semantic roles in Fluid Construction Grammar. In A. D. Smith, K. Smith, & R. Ferrer i Cancho (Eds.), *The evolution of language. proceedings of the 7th international conference (evolang 7)* (pp. 346–353). Singapore: World Scientific Press.