ISSN 2385-4138 (digital)
https://doi.org/10.5565/rev/isogloss.

Doing Romance linguistics in the 2020s

Roberta D'Alessandro
Utrecht University
r.dalessandro@uu.nl

How to cite: D'Alessandro, Roberta. 2023. Doing Romance linguistics in the 2020s. In *Romance grammars, context and contact,* eds. Alice Corr & Norma Schifano. Special issue of *Isogloss. Open Journal of Romance Linguistics* 9(2)/, 1-12.

1. Linguistic invariance and language variation

One of the challenges of generative syntax is how to handle variation. As a generativist, I try to look for uniformity across syntactic phenomena, so as to be able to draw some generalizations and ultimately to understand the laws governing certain language phenomena.

My linguistic career started with the study of impersonal *si* constructions in Italian. Italian being standardized, and impersonal *si* seeming to be fairly uniform in its use I expected to be able to identify some neat rules governing its distribution and its syntax, particularly with regard to agreement and interpretation. This assumption turned out to be completely on the wrong track. I was immediately confronted with a heterogenous set of data; the informants were all speakers of standard Italian, but they all used and interpreted impersonal *si* slightly differently from each other.

From that moment on, the idea of being able to find "the system" underlying the syntax of a specific language, let alone all languages, started to become somewhat shaky. Still, some generalizations are possible; some factors can be identified that have an impact on the syntax of some elements. In the case of *si*, it became fairly clear that its interpretation is dependent on whether the

event that the sentence denotes is bounded or not; on what the class of the verb is; and on which kind of adverbs are present in the clause (Cinque 1988, D'Alessandro 2004, 2004a,b, 2007). In addition, a number of other, grammar-external factors, turned out to be crucial: the interpretation and the agreement patterns of these "standard Italian" constructions had a neat geographical distribution, with some areas using *si* systematically as an impersonal pronoun only, others using it as a 1st plural personal pronoun (like French *on*), others allowing the existential reading, others only allowing the generic reading.

Variation, as every linguist knows, is the result of the interaction of many grammar-internal and grammar-external factors. Grammar-external factors, such as the geographic distribution of a phenomenon, or its standardization, are harder to account for. Standardized varieties tend to trigger sharper judgments on the part of the speakers, but the question is always, then: is this really their grammar or are they trying to "follow the rules" as they learnt them at school? Take for instance the Italian sentence in (1). The sentence is supposed to be wrong, every Italian grammar calls it out, and teachers at school correct children who say it. Yet, speakers continue to produce it, suggesting that it is a possible construction in Italian. As a linguist, how can you make sure than when you elicit data on sentences like (1) you are actually getting the speaker's intuitions and not the teacher's prescription?

(1) Italian

A me mi	piace	la	cioccolata
to me me.DAT	likes	the	chocolate

'I like chocolate'

Moving on to the study of dialects or non-standardized varieties of course amplifies the problem. If speakers do not have a reference standard, if they do not study a language systematically, if they use it only in a restricted environment and never write it, variation is bound to be pervasive. Does this mean that we need to abandon the idea of finding generalizations and universals? Not at all. In what follows I will try to show how looking at microvariation actually helps to understand language, by offering a privileged viewpoint on sentence structure.

2. Variation and microvariation

Syntax is about structure, the structure of sentences. The methodology of generative syntax is, at least in its intentions, scientific in the classical sense, which means: observing the data, formulating a hypothesis, setting up an experiment or collecting more data to prove or disprove the hypothesis, and developing a theory based on this. Developing a theory means making predictions on what will or will not be found in the data, making statements about what can never happen, and what will happen, given certain conditions.

We all know how messy linguistic data are. We also know, though, that this should not be an excuse to reduce complexity to the statement "anything goes". Not everything goes, not even in the physical world. If you let an apple fall from your hand, or from a tree, the apple will be attracted to the ground.

There are a couple of places on Earth where gravity seems to move the body upwards rather than downwards. There are some roads, in India, in Canada, in Scotland, where if you put the car in neutral it will move upwards, not downwards. Does this mean that gravity does not exist? Of course not. One needs to look very carefully at the conditions that alter the effects of gravity and at "what's wrong" with these particular places. We need to look at the "fringe" conditions, at the unusual contexts in which some exceptions emerge. The deviation tells us more than the well-behaved phenomena, but it doesn't immediately falsify the theory. And most importantly it doesn't make trying to build a theory a worthless enterprise. In other words: the fact that something is difficult, or seemingly impossible, shouldn't stop us from trying.

A theory makes predictions. Syntactic theory makes predictions on what we can find and what we can never find in a language. This is quite a powerful tool, if we want to understand the world. We do not just want to account for a set of phenomena; a set of observations is only the first step towards understanding language, and the world in general: we want to understand its laws, what governs the universe. There is no difference between a law of physics and a law of syntax: they apply to all relevant elements giving the same output. Like the laws of physics, the laws of syntax are blind. Like the laws of physics, the laws of language are shaded by other, external phenomena.

Theories are usually much harder to falsify than single hypotheses concerning one phenomenon, because they are the result of the convergence of several different matching hypotheses. One crucial aspect in which generative syntax differs from many other approaches is the emphasis that it places on predictions. If A then B. Romance languages are particularly useful when it comes to identifying laws and testing predictions: they have a long documented history, attestations that date back almost to their origin; they are fragmented, there are many of them, and they are rather widespread worldwide, which means that they are in contact with almost every other existing language family.

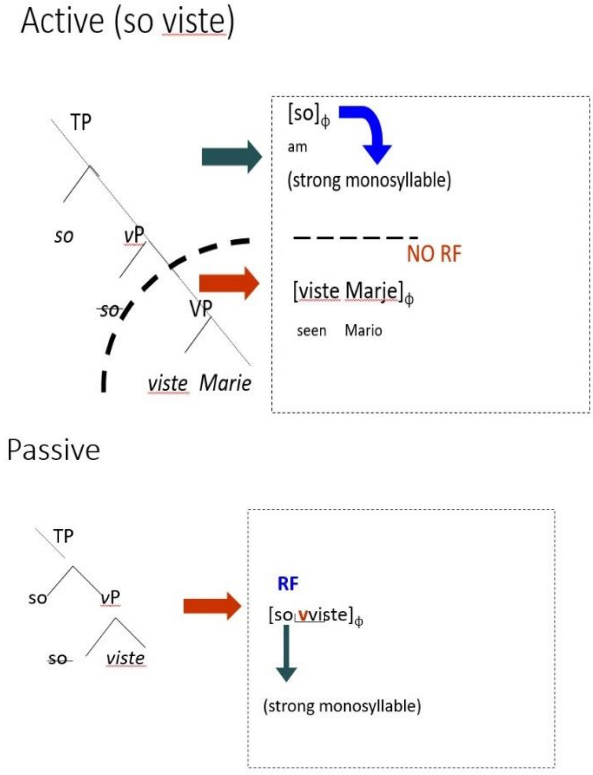
Romance languages are therefore ideal for testing hypotheses, and sometimes data from minority languages turn out to be crucial in showing that predictions made by the model were right. As an example, consider phase theory (Chomsky 2001); phases are domains of computation, which are postulated based on some observations regarding cyclic *wh*-movement, for instance. Chomsky (2001) proposed the *Phase Impenetrability Condition* (PIC), regulating transfer of linguistic material from Narrow Syntax to the other grammar modules, PF(Sensory-Motor system) and LF (Conceptual-Intentional system). The PIC identifies a specific point at which syntactic material is sent to the interfaces. The problem is that this sounds very much like a postulation: there seems to be no language that shows any marking for this PIC-defined domain. In fact, the study of Eastern Abruzzese, an upper-southern Italian variety spoken in the Abruzzo region of Italy, seems to offer evidence for the existence of the PIC-defined domains. In a number of studies with Theresa Biberauer (Biberauer & D'Alessandro 2006) and Tobias Scheer (D'Alessandro & Scheer 2013, 2015) it is shown that Eastern Abruzzese marks PIC domains and phases in two ways: by means of dedicated markers, as the doubled complementizer *ca* ('that') in (2) and more indirectly by means of morpho-syntactic phenomena that occur at PIC-borders like phonosyntactic doubling (*rafforzamento fonosintattico*), a sandhi

phenomenon that takes place between two words if certain phonological conditions are met, but more importantly if the words belong to the same domain. D'Alessandro & Scheer show that this phonological domain maps directly into the PIC domain (for a discussion of direct mapping, see D'Alessandro & Scheer 2015), as illustrated in example (3). These phenomena are exemplified in Figure 1, where the phonological rule of gemination is shown to apply to elements belonging to the same PIC-defined domain, i.e. in passives (*so**vv**iste*, 'I am seen') but not in actives (*so viste*, 'I have seen').

(2) Eastern Abruzzese, D'Alessandro & Scheer (2013:12)
Ca Màrijə **ca** li so vistə
that Mario that him am seen
'Mario, I saw him'

(3) Eastern Abruzzese, Biberauer & D'Alessandro (2006:1)
a. so vistə
am seen
'I have seen'
b. so vvistə
am seen
'I am seen'

Figure 1. Active and passive in Eastern Abruzzese and phase-based mapping



According to Biberauer & D'Alessandro (2006) and D'Alessandro & Scheer (2015), the mapping between syntax and the phonological component of grammar (PF) is phase-driven: if the elements that are sent from syntax to PF belong to the same domain, phonological phenomena such as *rafforzamento fonosintattico* can take place. Figure 1 illustrates that the PIC-induced domains, which are visible in the active part, as *v* is a phase head, but not in the passive, are mirrored in the phonology. The data presented here, which seem rather marginal and typologically rare, provide evidence for the existence of these domains in a way that other, larger languages, cannot. Minority languages prove therefore to be very important for syntactic theory.

3. Microcontact

Contact studies are almost never tackled from a theoretical/generative perspective. There are several reasons for this. First and foremost, language change in contact is largely the result of the interplay between grammar-internal and grammar-external factors. Grammar-external factors, such as the attitude of the speaker towards the contact language, the status of the language, the situation in which the speakers come into contact with the other language (whether they are bilingual from birth or if they start learning the other language, with or without institutional support), the so-called age of onset of the language all play a very big role in determining the direction that a language will take in contact (for an overview of these factors, see for instance Aikhenvald & Dixon 2007). These topics have been very prominent in the research on both contact between languages spoken in the same territory, such as Spanish and Basque, and heritage languages, meaning the languages spoken by the children of emigrants.

Conversely, syntactic theory has started to look at change, both diachronic and in contact, very late. However, no generativist will deny that external factors play a role: they are just very difficult to factor into a theoretical model as they are largely unpredictable and dependent on the socio-historical conditions in which contact takes place, as well as the personal trajectory of the speaker. The study of change in contact involving languages that are genetically related, part of the same family, is even more difficult, as it is almost never possible to ascertain whether the change we see in one or both languages is due to contact or to a normal evolution of the languages. The following quote by Aikhenvald (2007:9) is fairly clear in this respect:

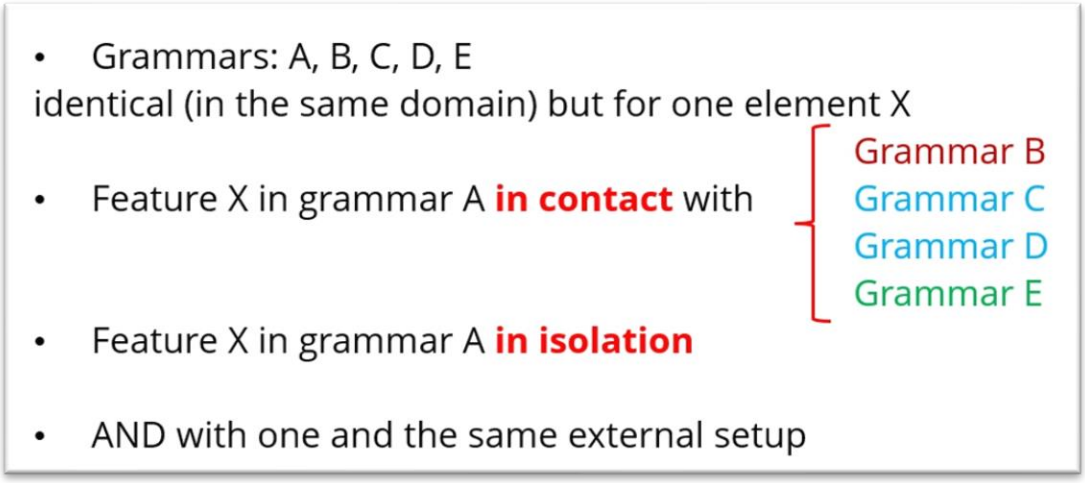
> "if languages are genetically related, we expect them to develop similar structures, no matter whether they are in contact or not. And if genetically related languages are in contact, trying to prove that a shared feature is contact-induced and not a chance result of Sapir's drift may be next to impossible"

However, there is a way to actually turn the apparent disadvantage of languages that are too similar into an advantage. The way to eliminate the problem of uncertainty is to use multiple cross-checking within a language family: we don't only compare language pairs, but we check one language against multiple almost

identical languages. This is the idea behind the *Microcontact* project, which was funded by the European Research Council and has the aim of understanding the mechanisms behind change in contact by examining closely related language varieties. The word *microcontact* was previously used, in sociolinguistics, to refer to situations of short contact between two or more languages, or to refer to contact between very small language communities (or even between two speakers only). The project gave this term a new definition (D'Alessandro 2015, 2018, 2021), to mean contact between minimally-differing varieties and cross-checking between them. One of the key methodological innovations of the microcontact enterprise has been that languages in contact are not considered in pairs, but in sets. Each variety, identical to all the others but for one feature, needs to be checked in contact with all other similar varieties.

Figure 2 offers an illustration of the *microcontact* methodology:

Figure 2 *Microcontact* methodology



Source: D'Alessandro (2021:7)

Considering five grammars that are identical in a given domain but for one element X, the element X will be checked in grammar A in contact with all other grammars. The contact grammars will ideally have different versions/values for the X element, which could be a syntactic feature or a syntactic configuration; the element X will also be checked in diachrony. In this way, it will be possible to examine the evolution of X and:

1. Ascertain whether X has undergone change in contact or just diachronic change
2. Identify the factors in the contact grammar that have influenced the change in feature X.

If the change has taken place along the same lines independently of the contact grammars, this must be attributed to a spontaneous, endogenous trigger. This can of course be checked against the diachronic data. Conversely, if the change has taken place in the direction of the contact grammars, we will be able to conclude that we are dealing with change in contact. Ideally, the grammars

should have similar values for the features in pairs, so that it will be possible to double-check the influence of the single factors.

To illustrate this methodology, let us consider the actual languages and phenomena investigated by *Microcontact*. The phenomena selected are: subject clitics and null subjects in contact; Differential Object Marking (DOM) in contact; and demonstratives and pronouns in contact.

The Italo-Romance varieties considered (after some were removed and others added in, depending on the availability of the data) were selected on the basis of the phenomena to be investigated. Every phenomenon had some core varieties that constituted the focus of the inquiry, and some "peripheral" varieties that were only considered as controls. The contact varieties selected were Argentinean Spanish, Brazilian Portuguese, Québecois French, and Italian, as well as English (as a control).

For subject clitics, the core varieties considered were Venetan and Friulian (and to a lesser extent Piedmontese). Both these varieties feature subject clitics, with slight differences in their systems. Heritage Venetan and Heritage Friulian are in contact with Spanish and Italian, two languages with no subject clitics; with Portuguese, a language with some instances of subject clitics/weak pronouns; with French, a language with subject clitics; and with English, a language with no subject clitics but with expletive subjects. Venetan and Friulian are also null-subject languages, as their clitics are agreement-like elements. Their null-subjecthood was checked in contact once again with Spanish and Italian, two null-subject languages; with Portuguese, a partial null-subject language; and only marginally with French and English (two non-null-subject languages).[1]

The study of DOM involved a larger sample of languages: northern Italian varieties like Friulian and Venetan, that do not feature DOM in the versions spoken in Italy; upper-southern varieties like Abruzzese, featuring person-oriented DOM in Italy, as well as Calabrian, featuring animacy and definiteness-driven DOM; and extreme southern varieties of Sicilian, also featuring DOM. These languages were observed in their heritage versions in contact with Argentinean Spanish (a widely-DOM variety, also featuring DOM with inanimates (Di Tullio, Saab & Zdrojewski 2019); Brazilian Portuguese, French and Italian (not featuring DOM).

The study of pronouns and demonstratives proved more complicated than the others, as the data elicitation with regard to demonstratives required extra fieldwork effort because most speakers had difficulty interpreting tests such as that featured in Figure 3, for deixis.

1 The reason why the data on contact with French are so few is that the Québecois speakers turned out to have a sociolinguistic profile that was somewhat different from that of the rest of our speakers; for that reason, the French contact study was moved to heritage speakers of Italo-Romance varieties in Belgium, whose profile is more in keeping with that of the rest of the speakers. The COVID-19 pandemic made it impossible, however, to complete the data collection in Belgium.

Figure 3. Elicitation test for demonstratives, *Microcontact* project

Figure 1: 'close to me' Figure 2: 'close to you' Figure 3: 'far from us'

Source: Andriani *et al.* (2022b:37)

Many speakers, especially those who were elderly and had a low level of education, had difficulty understanding for instance that the speech bubble represented something that the speaker was saying. The fieldworker had to resort to actual objects in the room to elicit the data. Because of the difficulty in data elicitation, the study of pronouns and demonstratives was carried out by examining some of the heritage data integrated with data from creole languages and earlier stages of the languages, where available.

The results of these studies, taken together, create an interesting picture of heritage microcontact, but crucially they show the importance of including the micro-dimension as well as cross-checking in the study of language contact. The results of our research are indeed often in contrast with what has been reported regarding heritage languages and contact.

With respect to DOM and null subjects, for instance, we see a different output in microcontact with respect to "macro"-contact, i.e. contact within language pairs made up of rather different languages. While contact between Italo-Romance varieties and English replicates the generalizations proposed for instance by Silva-Corvalán (1994); Luján & Parodi (1996); Montrul (2004); Montrul & Bowles (2009); Montrul & Sánchez-Walker (2013); Montrul, Bhatt & Girju (2015) and others regarding the weakening of these marked constructions, the data collected on microcontact go in a totally different direction (Andriani *et al*. 2022a, b).

This is possibly the result of several factors: first, if the speaker is unaware of the locus of variation and can't identify the exact difference between two minimally-differing phrases, they will resort to "cognitive" methods to resolve the conflict between the grammars. For instance, when two minimally-differing varieties come into contact, one of the first phrases that the speakers identify is the topic/aboutness phrase. They usually mark that with a DOM-marker, which they perceive as something related to the saliency of the object. In languages that are different lexically, structurally, and maybe typologically, such as English and Spanish, the speakers do manage to identify the locus of variation within a VP; for instance, they recognize the object in English and they see that it is never marked. In this case, they resort to the strategies that have been described for language contact, namely they usually select the unmarked version of a given construction (a transitive verb in this case), as that is the simplest to handle. In those cases in which dialects of the same macrovariety are in (micro-)contact, the speakers do not seem to mind an increase in complexity (see Andriani *et al*. 2022a).

Regarding the issue of whether change in contact is accelerated diachronic

change, the answer seems twofold: when φ-features, like person and number, are involved, this does indeed seem to be the case (see Terenghi 2021, 2022); when, however, we are dealing with discourse features, the direction of change seems more unpredictable, and the interspeaker variation much greater (D'Alessandro, Putnam & Terenghi 2022).

The generalizations regarding the data collected can be found in two articles (Andriani *et al*. 2022a,b), while Andriani *et al.* (2022b) offers a clear description of the methodology of the data collection.

4. Data collection

Collecting the data for a project with such a large empirical basis is no easy enterprise. We organized the research into 3 phases: a first phase, aimed at identifying speakers with the right profile, was carried out using an interactive atlas for data crowdsourcing (https://microcontact.hum.uu.nl/). Computer-proficient, usually young, speakers were contacted through several institutions and asked to record their parents and grandparents and to upload the recording onto the atlas. The recordings were accompanied by basic sociolinguistic information, such as age, place of provenance, languages spoken, level of education, etc. The recordings were then coded for language and published on the atlas.

While many Italians responded to the call, we received very few recordings from the Americas; the recordings that are featured on the atlas were collected by our fieldworkers.

The second part of the data collection was carried out through fieldwork: our researchers went to Canada, the US, Argentina and Brazil, to interview the speakers of Italo-Romance. While our intention was to record entire families, so as to have as accurate an idea as possible of the input that the 2nd and 3rd generation received, we very rarely found entire families (grandparents, parents, and grandchildren) speaking to each other in Italo-Romance. The most common situation that we encountered was that of grandparents using their language with their grandchildren, while parents used the contact/country language with them (see Andriani et al 2022b).

The second and longer fieldwork period was planned to take place in the spring of 2020, and was cancelled due to the COVID-19 pandemic. Many of the tests we had in mind had to be restructured or canceled. We collected some data online, but the cancelation of the second fieldwork period was a huge drawback for the project.

The routine followed for the data collection included a first part of free speech, with the aim of putting the speakers at ease with the target language; a HALA test (O'Grady *et al*., 2009) to check fluency in the target language; some forced choice tests; some sentence completion tasks; and some repetition tasks. The full set of tests is described in Andriani *et al*. (2022b).

While *Microcontact* is a theoretical project, collecting data from Italian emigrant communities was also a very interesting sociolinguistic exercise. To begin with, every community has a different profile. As an example, speakers who moved to remote parts of Brazil had a much greater possibility of continuing to

speak their own language than those who moved to the big cities, who were forced by law to abandon their native language. Speakers who moved to Argentina started using Spanish much earlier than the speakers who moved to Brazil started using Portuguese, possibly because of language proximity. The varieties spoken in New York, like those spoken in other big cities, have a completely different profile with respect to those spoken in smaller centers. In NYC, where large communities of Italians came together from many different regions, the language spoken now is a sort of mixed Italo-Romance (see for instance Andriani *et al*. 2022b), while the original varieties have been almost completely lost.

Any linguistic project requires attention to the data and a careful method for data collection. Without that part, no linguistic research can be effective. Doing Romance linguistics, in this respect, is no different than doing Germanic, or Indo-Aryan, linguistics. It is enriching, challenging and fulfilling, in a way that only linguistics research can be.

References

Aikhenvald, Alexandra. 2007. Grammars in contact: a cross-linguistic perspective. In A. Aikhenvald & R.W.M. Dixon. 2007 (eds), 1-66.

Aikhenvald, Alexandra & R.M.W. Dixon (eds). 2007. *Grammars in Contact*: *A cross-linguistic typology*. Oxford/New York: Oxford University Press.

Andriani, Luigi, D'Alessandro, Roberta, Frasson, Alberto, Van Osch, Brechje, Sorgini, Luana & Silvia Terenghi. 2022a. Adding the microdimension to the study of language change in contact. Three case studies. *Glossa* 7(1), https://doi.org/10.16995/glossa.5748

Andriani, Luigi, Casalicchio, Jan, Ciconte, Francesco Maria, D'Alessandro, Roberta, Frasson, Alberto, Van Osch, Brechje, Sorgini, Luana & Silvia Terenghi. 2022b. Documenting Italo-Romance minority languages in the Americas. Problems and tentative solutions. In: Nevins, A. & M. Cole (eds), *Contemporary research in minority and diaspora languages of Europe*, Language Science Press (online first).

Biberauer, Theresa & Roberta D'Alessandro. 2006. Syntactic doubling and the encoding of Voice in Abruzzese. In D. Baumer, D. Montero & M. Scanlon (eds), *Proceedings of the 25th West Coast Conference on Formal Linguistics*, , 87-95. Somerville, MA: Cascadilla Proceedings Project.

Chomsky, Noam. 2001. Derivation by phase. In M. Kenstowicz, (ed.), *Ken Hale: A Life in Language* , 1–52. Cambridge, MA: MIT Press.

Cinque, Guglielmo. 1988. On *si* constructions and the theory of ARB. *Linguistic Inquiry* 19: 521–81.

D'Alessandro, Roberta. 2004a. Impersonal *si* constructions: how semantics determines agreement. *Proceedings of NELS 34.*

D'Alessandro, Roberta. 2004b. *Impersonal si constructions. Agreement and Interpretation.* PhD dissertation, University of Stuttgart.

D'Alessandro, Roberta. 2007. *Impersonal* si *constructions.* Berlin/New York: Mouton de Gruyter.

D'Alessandro, Roberta. 2015. *Microcontact. Language variation and change from the Italian heritage perspective.* ERC-CoG project.

D'Alessandro, Roberta. 2018. Microcontact: How to take synchrony, diachrony, and sociolinguistics into account when studying variation. Keynote talk, *Transdisciplinary Approaches to Language Variation*. Tromsoe, 19-20 April 2018.

D'Alessandro, Roberta. 2021. Syntactic change in contact: Romance. *Annual Review of Linguistics* 7(7), 309:328.

D'Alessandro, Roberta, Putnam, Michael T. & Silvia Terenghi, 2022. On the similarities and differences between diachronic & contact-induced syntactic change. Paper presented at GLOW in Asia XIII. Hong Kong, 6 August 2022.

D'Alessandro, Roberta & Tobias Scheer. 2013. Phase-head marking. In: Hisao Tokizaki and Yoshi Dobashi, eds, Universal Syntax and Parametric Phonology. Special Issue of Linguistic Analysis 38:4, 305-330.

D'Alessandro, Roberta & Tobias Scheer. Modular PIC. *Linguistic Inquiry* 46:4, 593-624.

Di Tullio, Ángela, Saab, Andrés & Pablo Zdrojewski. 2019. Clitic doubling in a doubling world. The case of Argentinean Spanish reconsidered. In Á. Gallego (ed.), *The Syntactic Variation of Spanish Dialects,* 213-242. Oxford/New York: Oxford University Press.

Luján, Marta & Claudia Parodi. 1996. Clitic-doubling and the acquisition of agreement in Spanish. In J. Gutiérrez-Rexach & L. Silva Villar (eds), *Perspectives on Spanish linguistics* (vol. 1), 119–138. Los Angeles, CA: UCLA, Department of Linguistics.

Montrul, Silvina. 2004. Subject and object expression in Spanish heritage speakers: A case of morphosyntactic convergence. *Bilingualism: Language and cognition* 7(2). 125–142. DOI: https:// doi.org/10.1017/S1366728904001464.
Montrul, Silvina, Bhatt, Rajesh, & Girju, Roxana. 2015. Differential Object Marking in Spanish, Hindi, and Romanian as Heritage Languages. *Language* 91(3). 564–610. DOI: https://doi. org/10.1353/lan.2015.0035.

Montrul, Silvina & Melissa Bowles. 2009. Back to basics: Differential Object Marking under incomplete acquisition in Spanish heritage speakers. *Bilingualism: Language and Cognition* 12(3). 363–383. DOI: https://doi.org/10.1017/S1366728909990071.

Montrul, Silvina & Noelia Sánchez-Walker. 2013. Differential Object Marking in child and adult Spanish heritage speakers. *Language Acquisition* 20(2). 109–132. DOI: https://doi.org/10.1080/ 10489223.2013.766741.

O'Grady, William, Schafer, Amy J., Perla, Jawee, Lee, On-Soon & Julia Wieting. 2009. *A psycholinguistic tool for the assessment of language loss: The HALA Project. Language Documentation & Conservation* 3. 100–112.

Silva-Corvalán, Carmen. 1994. *Language contact and change: Spanish in Los Angeles*. Oxford: Clarendon.

Terenghi, Silvia. 2021. Stable and unstable person features: A structural account. In A. Farinella & A. Hill (eds), *NELS 51: Proceedings of the Fifty-First Annual Meeting of the North East Linguistic Society*, vol. 2, 229–242. Amherst, MA: University of Massachusetts Graduate Linguistics Student Association.

Terenghi, Silvia. 2022. *Missing Person: Structure and change in Romance demonstratives*. PhD dissertation, Utrecht University.