

Intervention effects in object relatives in English and Italian: a study in quantitative computational syntax

Giuseppe Samo

Department of Linguistics
Beijing Language and Culture University
samo@blcu.edu.cn

Paola Merlo

Department of Linguistics
University of Geneva
Paola.Merlo@unige.ch

Abstract

Discontinuous dependencies are one of the hallmarks of human languages. The investigation of the locality constraints imposed on such long-distance dependencies is a core aspect of syntactic explanations. The aim of this work is to investigate locality constraints in object relative clauses adopting a theory-driven and quantitative point of view. Based on a comparison of the theoretically expected and the observed counts of features of object relative clauses, we study which set of features plays a role in the syntactic computation of locality (type, number, animacy). We find both effects predicted by a narrow and a broad view of intervention locality. For example, in Italian the feature number triggers a numerically stronger effect than in English, a prediction of the narrow, grammar-driven view of locality. We also find that the feature animacy plays a role in the frequency of object relative clauses, an effect predicted by a broader view of locality.

1 Introduction

The aim of this work is to investigate locality issues adopting a quantitative computational syntax point of view (Merlo, 2016). Quantitative computational syntax uses large-scale resources and simple computational models in order to answer quantitative linguistic questions. In this paper, we concentrate on the quantitative aspect of long-distance dependencies according to a theory of intervention. Based on a comparison of the theoretically expected and the observed counts of features in grammatical structures, we study which set of features plays a role in the syntactic computation of long-distance dependencies.

A core distinguishing property of human languages is the ability to interpret discontinuous elements as if they were a single element. Sometimes these elements are distant in the string. These are called long-distance dependencies. For example, sentence (1a) is an object-oriented restrictive relative clause, where the object of the verb *show* is also the semantic object of the verb *wash*, connecting two distant elements. The sentence (1b) is a subject-oriented restrictive relative clause, where the semantic object of the verb *show* is also the subject of the verb *wash*.¹

(1a) Show me the elephant that **the lion** is washing <the elephant>.

(1b) Show me the elephant that <the elephant> is washing the lion.

Long-distance dependencies are not all equally acceptable (Rizzi, 2004). The facts involving them are complex, and a precise description encompassing all phenomena is one of the major topics of research in current linguistic theory (Rizzi, 1990; Gibson, 1998). We study here the predictions of an intervention theory of locality (Rizzi, 1990). In a nutshell, a long-distance dependency between two elements in a sentence is difficult, and often impossible in child grammar (Friedmann et al., 2009), if a *similar* element intervenes. For example, sentence (1a) causes trouble for children while (1b) does not, because in (1a) the *lion* intervenes between the two discontinuous occurrences of *elephant* (one pronounced one silent), while in (1b) there is no intervener.

Core to the explanation of these facts is the notion of *intervener*. An intervener is an element that is *similar* to the two elements that are in a long-distance relation, and structurally intervenes between the two, blocking the relation. In our examples, the intervener is *the lion*, shown in bold.

¹The unpronounced element(s) in the long-distance relation are indicated by < >.

Notice that here and in all the following, intervention is defined structurally and not linearly. Linear intervention that does not structurally hierarchically dominate (technically *c*-command) does not matter, as shown by the contrast **When do you wonder who won?/You wonder who won at five* compared to *When did the uncertainty about who won dissolve?/The uncertainty about who won dissolved at five* (Rizzi, 2013). Notice also that the non *c*-commanding, more acceptable alternative creates a linearly longer dependency than the *c*-commanding more difficult one, therefore also showing that length of the dependency does not directly affect acceptability.

Defining and justifying which properties come into play in computing whether two elements are similar or not is therefore a crucial element in this explanation. In this paper, we briefly review some results from the theoretical and experimental literature that have attempted to characterise precisely this notion of intervener in the case of object relative clauses. Based on their findings, we develop hypotheses of the expected corpus distributions.

2 Object relatives and intervention locality

A robust set of experimental studies and results on both production and comprehension of relatives clauses, both subject relatives and object relatives on acquisition (Friedmann and Novogrodsky, 2004), on adult processing (Frauenfelder et al., 1980), and on pathology (Grillo, 2008) confirms that object relatives are harder than subject relatives, in various respects both in children and adult grammar, as shown in example (1) above. The intervention locality explanation ascribes this difficulty to the fact that the subject acts as intervener between the head of the relative clause and the object position in object relatives, while in subject relative clauses no intervention occurs (Friedmann et al., 2009).

According to this theory, the crucial property in intervention is not the amount of material that can be considered as intervener, but rather its quality. If the head of the relative clause and the intervener share some computationally relevant features, this leads to slower processing for adults.

One important aspect in verifying the intervention-based explanation of the difficulty of object relatives, then, is determining which features trigger the intervention. Object relatives are grammatical structures and thus the type of intervention could be qualitatively different from other long-distance constructions creating ungrammatical sentences, such as long distance complex questions (*wh*-islands). According to recent studies (Belletti et al., 2012), the relevant features in intervention in object relative clauses are those features which could be considered syntactically relevant in the language. In particular, those features able to trigger the movement of syntactic elements such as the subject and the verb. In English and Italian, features such as *number* and *person* (Bentea, 2016) have been investigated and the various forms of noun phrases, such as pronouns (head) vs. maximal projection, indicated as XP, (we will call it *type*) (Friedmann et al., 2009).² Finally, the status of an *animacy* feature remains controversial; some results argue in favour of an ameliorative effect (Brandt et al., 2009), some suggest animacy has no effect (Adani, 2012). Some recent studies show a clear effect of animacy as an intervention feature in *wh*-islands, another kind of long-distance dependency (Villata and Franck, 2016). Further evidence for the need for a finer theory of locality comes from studies in language pathology and language acquisition, where, within the same language, the grammar of different populations (e.g. the grammar of adults vs. the grammar of children) exhibits different locality effects (Grillo, 2008; Friedmann et al., 2009; Belletti et al., 2012).

3 Quantifying the hypotheses

We choose to investigate the features of *type*, *number* and *animacy*. We show in Table 1 some examples of relatives clauses with these features. We select these features to explore several dimensions of variation. First of all, the notion of *type* (head or maximal projection) goes back to the core formulation of

²Several pieces of work in language acquisition, adult processing and language pathology have investigated a set of morphosyntactic features such as *number*, *animacy*, *gender*, *case* and *lexical restriction*. A non-exhaustive list of reference is to be found at the ERC Syncart website <https://www.unige.ch/lettres/linguistique/syncart/cartographylocality/references/thematic-order/th/> edited by Karen Martini. According to some proposals these sets of features may be organised in a structural typology expressed as morphosyntactic features (Rizzi, 2004). Argumental: *person*, *number*, *gender*, *case*; Quantificational: *Wh*, *Neg*, *measure*, *focus*; Modifier: *evaluative*, *epistemic*, *Neg*, *frequentative*, *celerative*, *measure*, *manner*; *Topic*.

head of relative		subject	
the debate	which	we	held
XP, singular, inanimate		head, plural, animate	
these lovely little chocolates	that	we	get
XP, plural, inanimate		head, plural, animate	
Il terreno	che	l' acqua	copre
<i>the ground</i>	<i>that</i>	<i>the water</i>	<i>covers</i>
XP, singular, inanimate		XP, singular, inanimate	

Table 1: Examples of object relative clauses in several featural configurations. The examples are naturally occurring clauses extracted from the Universal Dependency corpora. XP=maximal projection.

intervention locality theory and has been shown to be active in the acquisition of object relative clauses (Rizzi, 1990; Friedmann et al., 2009). The morphosyntactic feature of number (singular or plural) has been studied because it is related to the richness of the verbal morphological system (Bentea, 2016). It has been argued that a rich morphological system triggers greater verb movement (Pollock, 1989). If this is the case, then morphosyntactic features in Italian may show a different strength of intervention than in English, since Italian has richer morphology and a greater movement of the verb. Animacy is a lexical semantic feature, whose influence on intervention is still controversial, as indicated above, and for which there is no reason to expect a cross-linguistic difference. Based on the findings in the theoretical and experimental literature, we can formulate the following questions.

1. Do the features *type*, *number* and *animacy* play a role in intervention effects?
2. If the features play a role in intervention effects, are these effects stronger in one of the two languages?

To answer these questions quantitatively based on corpus counts, we need to define the concept of similarity, central to the notion of intervention, and a linking hypothesis.

- *Similarity* The head of the relative clause and the intervener are *similar* if their features match.
- *Feature match* A *feature match*, $match_f(C, I)$, is true iff, for a given feature f , the head of the relative C and the intervener I have the same value.
- *Linking hypothesis* If a feature is a stronger intervener, we expect it to create greater inacceptability and hence surface less often in a corpus in a match configuration.

In this work, we make use of observational data provided by corpora, and operate on counts. We will refer to the notion of *observed counts*, as usual, as the counts in the corpus, and to *expected counts* as the counts of the features that we would expect based on their distribution in a setting where intervention is not at play and, therefore, they do not interact with each other. That is, the expected counts are the counts we would expect given the probability of the two features to cooccur independently of intervention, proportionally to the size of the corpus. Specifically, an object-oriented relative clause brings into play the object of the verb and its features, the noun phrase that is being relativized, and the subject of the sentence and its features, the intervener. Precisely, let C_s^f be the counts of a subjects feature and C_o^f be the counts of an object feature in a sample of size S . Let T be the total number of observations. Then, the expected counts of subject and object features occurring in a sentence with intervention are calculated as $C_s^f/S \times C_o^f/S \times T$, namely the product of the relative frequencies of these two elements, counted independently, in a sample, scaled by the total size of the corpus.

As we said at the beginning, we use corpus counts and frequencies in the spirit of the computational quantitative syntax framework: *differentials in counts are the expression of underlying grammatical properties*. In this respect, our quantitative hypotheses below are to be contrasted to an H_0 hypothesis that would predict that grammatical properties are uncorrelated to observed counts in a corpus, because corpus counts are effects of usage, while grammar makes no predictions about them, and as such there is

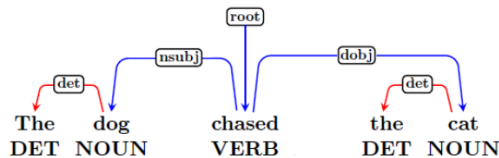


Figure 1: Canonical order

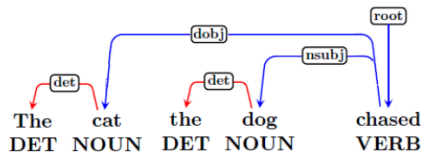


Figure 2: Non-canonical order

no expectation of distribution of counts beyond the observed ones. Based on these notions, we formulate then the following quantitative hypotheses.

H_1 Both in Italian and English, if the features *type*, *number* or *animacy* trigger intervention effects, we expect match configurations to be less frequent than expected. (Possibly, non-match configurations are more frequent than expected.)

H'_1 If the features *number* triggers intervention effects, the effect (the difference between expected and observed matches) should be larger in Italian than in English.

4 Materials and methods

Our hypotheses above follow a common schema that requires calculating the observed counts of a feature in the corpus and compare it to the counts we would expect if intervention was not at play. The annotated corpora we use are the universal dependency treebanks for Italian and English.

4.1 The corpus

We extract our counts from the Italian and English Universal Dependencies (UD) treebanks (Nivre, 2015) version 2.0 (<http://universaldependencies.org/>).³ The data comes from five different treebanks: English ParTut (Bosco and Sanguinetti, 2014), English LinEs (Ahrenberg, 2015), English UD (Bies et al., 2012), Italian ParTut (Bosco and Sanguinetti, 2014), Italian UD (Bosco et al., 2013). They comprise a variety of text genres. For English: blogs, social media, reviews, fiction, nonfiction, spoken legal, news, wiki For Italian : legal, news, wiki. In what follows, the analysis of the phenomenon will not be presented according to the different corpora, but all the different treebanks for each languages will be merged and thus we will refer to Italian and English data without the specification of the treebanks. The reason for this merge is the observation that different corpora show very large fluctuations in distributions of grammatical constructions, in general, and specifically for relative clauses (Roland et al., 2007; Belletti and Chesi, 2014, 3-5). By merging the different corpora, we limit the corpus-related biases in our counts and increase the generality of our results.

We choose UD because a comparative investigation is made possible by the uniformity of UD annotation across languages, although comparative syntax investigations are not the purported goal for which Universal Dependencies was developed.⁴ Our theory is formulated in terms of discontinuous dependents, which are not explicitly encoded in UD, so we need to adapt our searches to a dependency annotation. UD dependencies can be defined by the syntactic relation label and its direction. Syntactic relations, in

³To extract the construction of interest here, object relative clauses, we used the tool SETS, a query UD treebank search tool developed and maintained by the University of Turku (Haverinen et al., 2013).

⁴Universal Dependencies (UD) aim to provide a cross-linguistically uniform syntactic representation to advance multilingual applications of natural languages processing (Nivre, 2015).

Treebank	objects	left objects	OR	%OR
English ParTut (Bosco and Sanguinetti, 2014)	3186	51	44	86
English LinEs (Ahrenberg, 2015)	5985	139	16	11
English UD (Bies et al., 2012)	15259	403	191	47
Italian ParTut (Bosco and Sanguinetti, 2014)	3142	56	49	71
Italian UD (Bosco et al., 2013)	14639	549	216	39

Table 2: Object dependencies to the left (left objects) and object relatives (OR) in English and Italian.

	English		Adjusted English		Italian		Adjusted Italian	
	Subject	Object	Subject	Object	Subject	Object	Subject	Object
XP	.49	.91	.49	1.0	.62	.86	.62	1.0
head	.48	.09	.48	.00	.05	.14	.05	.00
null	.03	-	.03	.00	.33	-	.33	.00
singular	.70	.73	.70	.73	.74	.67	.74	.67
plural	.30	.27	.30	.27	.26	.33	.26	.33
animate	.93	.22	.93	.22	.78	.20	.78	.20
inanimate	.07	.78	.07	.78	.22	.80	.22	.80

Table 3: The proportions of expected counts and adjusted expected counts of the different features, in English and Italian, for subjects and objects.

a given language, have a canonical direction, for example subjects are left directed arcs and objects right directed arcs in an SVO language as shown in Figure 1. Non-canonical directions, left for object, for example, can be indicators of discontinuous dependency, as illustrated in Figure 2. The query search ‘VERB > obj@L_’ provides all the occurrences of the objects on the left of the verb. Statistics on the collected constructions and the number of object relatives after manual inspection and validation are shown in Table 2. This table shows that the occurrences of objects whose dependency is on the left represent a very small proportion in corpora; a sizable proportion of these left dislocated objects are object relatives (ORs).⁵

4.2 The counts

Like many experimental settings, our predictions are formulated in terms of differences in numerical observations. In this case, we study the difference between the expected counts and the observed counts.

Expected counts To validate our hypotheses, we establish the expected counts of morpho-syntactic and animacy features in all syntactic configurations. We sampled randomly one-hundred sentences in the English UD treebanks and one-hundred sentences in the Italian UD treebanks, selected with the SETS treebank search tool (the search query ‘_ <obj _’ provided all the results of sentences having an overt object). We then coded the features of subjects and objects. The subjects in a normal construction become the interveners in an object relative clause and the objects are the element that undergoes the relative clause long-distance dependency. Post-verbal subjects in Italian were considered as having the same intervening effects as preverbal subjects and were included in the subject counts.

The three features we code are *type*, *number* and *animacy*. Number is indicated in the morphology of the noun phrase and agreement in the sentence and leads to unambiguous classification. Type and animacy require judgement. Type can have the values head or maximal projection: pronouns were coded as heads, other noun phrases were coded as maximal projections. For animacy, elements such as *the parliament*, *the commission* (animate collectives) were labelled as animate elements. Those relative heads and subjects involving human beings, animals or groups of human beings and/or animals (e.g. the

⁵Other than object relatives, left dislocated objects have been analysed as Topics, anaphoric clitics, resumptive clitics, Wh-elements, Embedded Wh-elements, Wh-DP and rare cases of imperfect annotations with heterogenous distributions in the different treebanks.

Match			Relative head			Intervener			Sentence
type	num	an	type	num	an	type	num	an	
0	0	0	XP	sg	in	head	pl	an	<i>the foreign investment</i> that they need to help their economies grow
0	1	0	XP	pl	in	head	pl	an	<i>the fees</i> that they charge
1	0	0	XP	sg	in	XP	pl	an	<i>a luxury</i> that only rich countries can afford
1	0	1	XP	sg	an	XP	pl	an	<i>a better person</i> that people are wanting to hire
1	1	0	XP	sg	in	XP	sg	an	<i>a realist technique</i> which French novelist Marcel Proust later named <i>retrospective illumination</i>
1	1	1	XP	sg	in	XP	sg	in	a format that Access recognizes

Table 4: Examples of OR clauses in several featural configurations in English. The examples show the values of the features and if they match (1) or not (0) between head of the relative clause (in italics) and the intervener (in bold). The examples are naturally occurring clauses extracted from the UD corpora.

Match			Relative head			Intervener			Sentence
type	num	an	type	num	an	type	num	an	
0	0	0	XP	pl	in	null	sg	an	<i>i luoghi</i> che [0] aveva visitato spesso (<i>the places that (s/he) had visited often</i>)
0	0	1	XP	pl	in	head	sg	in	<i>i seri problemi</i> che ciò genera (<i>the serious problems that this engenders</i>)
0	1	0	XP	pl	in	null	pl	an	<i>la prima cosa</i> che [0] vide (<i>the first thing that (s/he) saw</i>)
0	1	1	XP	sg	an	null	sg	an	<i>l'associazione per l'abolizione della pena di morte</i> che [0] aveva fondato (<i>the association for the abolition of the death penalty that (s/he) had founded</i>)
1	0	0	XP	pl	in	XP	sg	an	<i>i sonetti</i> che Shakespeare intendeva pubblicare (<i>the sonets that Shakespeare meant to publish</i>)
1	0	1	XP	pl	in	XP	sg	in	<i>le limitazioni</i> che la legge stabilisce (<i>the limitations that the law dictates</i>)
1	1	0	XP	sg	in	XP	sg	an	<i>il tipo di effetto</i> che Balzac tentava di ottenere nelle sue opere (<i>the type of effect that Balzac attempted to obtain in his works</i>)
1	1	1	XP	sg	in	XP	sg	in	Il terreno che l' acqua copre (<i>the ground that the water covers</i>)

Table 5: Examples of OR clauses in several featural configurations in Italian. The examples show the values of the features and if they match (1) or not (0) between head of the relative clause (in italics) and the intervener (in bold). Empty subjects are indicated as [0]. The examples are naturally occurring clauses extracted from the UD corpora. A literal translation of the Italian clauses is given in parentheses.

government, the European Union, Russia), were labelled as animate. The nouns denoting non-human beings and non-animals were considered inanimate.

The expected counts of the different features, in English and Italian, for subjects and objects are shown in the contingency Table 3. We include a column indicating adjusted counts. This is based on the observation that relatives clauses with a pronoun head or a null head are extremely rare or impossible. This is because neither English nor Italian are null-object languages and, thus, a null relative head will result in an ungrammatical sentence. So, in fact, the counts of these features in a relative clause are different from their distribution in a simple transitive sentence. We will use the adjusted expected counts for our comparisons.

Observed counts We also need to collect the observed counts of cooccurrence of features. The statistics of observed relative clauses are given in Table 2, which indicates that we have 251 object relative clauses for English and 265 for Italian. A manual analysis by the first author coded the features and the match vs. mismatch conditions. We show some examples of relatives clauses with the feature coding in Tables 4 and 5.⁶ The Boolean values indicate the feature coding in a summarised way, by indicating if the features of the head of the relative clause and of the intervener match (1) or not (0). This encoding is just a shorthand for illustratory purposes. It was not added to the coding of features.

For example, the sentences *the foreign investment that they need to help their economies grow* is coded $\langle 0,0,0 \rangle$, as none of the features type, number, and animacy match, as *the foreign investment* is of type maximal projection, number singular, and inanimate, while *they* is a plural, animate pronoun head. The example *the fees that they charge* is coded as $\langle 0,1,0 \rangle$, as *the fees* is a plural, inanimate maximal projection, while *they* is a plural, animate pronoun head. Finally, the example *Il terreno che l'acqua copre (the ground that the water covers)* is coded as $\langle 1,1,1 \rangle$, as both *il terreno (the ground)* and *l'acqua (the water)* are singular, inanimate maximal projections.

5 Results and discussion

The calculations of expected counts and actual observed counts, the probabilities of these observations under a binomial distribution and their statistical significance are shown in Table 6. The binomial test gives us the probability of k successes in n independent trials, given a base probability p of an event. The event in our case is the cooccurrence of two features. So, for example, the binomial distribution tells us the probability of the (anim, anim) pair of features in English. Specifically, it tells us the probability of 20 successes in 251 trials given a base probability of the event of $.93 \times .22 = 0.2046$. The base probability of the event is, in our case, the product of the probabilities of the subject and object features, that is the probability of cooccurrence of these two features if they were independent and not in an intervention configuration. If certain conditions are met, the binomial distribution can be approximated by the normal distribution and a significance test can be performed. We calculate the cumulative probability distribution: the probability that the observed counts are exactly as observed, or greater, if the observed counts are larger than the expected counts, or the probability that the observed counts are exactly as observed, or smaller, if the observed counts are smaller than the expected counts. The z -score gives us the (one-tailed) probability of exactly, or greater/smaller counts than the expected counts.

The results that confirm the hypotheses, because they are in the right direction numerically and statistically significant, are shown in bold. These results are mixed, but have some interesting sub-regularities. In the match configurations, hypothesis H_1 is confirmed for the features *type* and *animacy* in most cases, for both English and Italian. Only the (inanimate, inanimate) pair in English is numerically smaller than expected, and as such confirming the hypothesis, but not significantly so statistically. For these features, we also observe an increase of observed non-match configurations, where statistically valid conclusions can be drawn. Mismatches are robustly more frequent than expected, especially in Italian. This is possibly compatible with an intervention effect, if we take these preferences for non-matching configurations as preference for alternative forms to avoid matches. We also observe that for these features, both in the match and mismatch configuration, the hypothesis is not confirmed only in the smaller or zero observed

⁶The supplementary materials with all the coded data are also available from the first author.

Match condition English						
HRel	Interv	Expected	Observed	p	Binomial p	$z-p$
XP	XP	123.0	108	0.490	0.033	0.033
sing	sing	128.7	132	0.511	0.341	0.341
plur	plur	20.3	22	0.081	0.382	0.393
anim	anim	51.4	20	0.205	0.000	< .000001
inan	inan	13.7	12	0.055	0.399	0.384

Match condition Italian						
HRel	Interv	Expected	Observed	p	Binomial p	$z-p$
XP	XP	164.3	149	0.62	0.0313	0.03053
sing	sing	131.4	138	0.496	0.218	0.218543
plur	plur	22.7	34	0.86	0.011	0.007814
anim	anim	41.3	23	0.156	0.0006	0.001263
inan	inan	46.6	27	0.176	0.0006	0.001009

Mismatch condition English						
HRel	Interv	Expected	Observed	p	Binomial p	$z-p$
XP	head	120.5	135	0.480	0.383	0.038
XP	null	7.5	0	0.030	0.0005	<i>n.v.</i>
sing	plur	47.4	49	0.219	0.203	0.202
plur	sing	53.2	40	0.189	0.131	0.132
anim	inan	3.9	0	0.015	0.022	<i>n.v.</i>
inan	anim	182.1	211	0.725	0.00001	0.00003

Mismatch condition Italian						
HRel	Interv	Expected	Observed	p	Binomial p	$z-p$
XP	head	13.3	29	0.050	0.000075	0.000009
XP	null	87.5	101	0.330	0.0453	0.044109
sing	plur	46.2	59	0.174	0.0249	0.022341
plur	sing	64.7	48	0.244	0.0088	0.010407
anim	inan	11.7	0	0.044	0.000007	0.000415
inan	anim	165.4	229	0.624	0.00000001	0.000001

Table 6: Expected counts and observed counts. Expected counts are based on adjusted proportions. English $N = 251$, Italian $N = 265$. p is the prior probability of the event. Binomial p indicates the probability of the observed counts under a binomial distribution (the binomial test). $z-p$ is the statistical significance of the binomial probability. *n.v.* indicates that conditions are not met for a valid calculation of statistical significance. The $z-p$ gives us the (one-tailed) probability of exactly the observed, or greater/smaller counts than the expected counts, for $\alpha = 0.5$. Results confirming the hypotheses are in bold.

counts. We reserve to investigate further if this result is due to a too small sample size. Notice that the feature *animacy* clearly triggers intervention effects, both in Italian and English, with a big preference for the mismatch configuration and dispreference for the match. This is quite interesting, as the results concerning this features are still not entirely converging. Experimental work on *wh*-islands indicate that the feature is relevant (Villata and Franck, 2016), showing a similar preference for mismatches over controls as what we found here, but results from acquisition seem to indicate it is not (Adani, 2012). Our results show that, at least in the adult grammar in written text, animacy makes a difference to the preference for choice of relative head and intervener in an object relative.

Instead, neither H_1 nor H'_1 are convincingly confirmed for the feature *number*. For H_1 , none of the predictions in the match configurations are confirmed and only half of the mismatch configurations are. With respect to the cross-linguistic hypothesis H'_1 about the feature *number*, the numerical differences do show a greater differential in Italian than in English, but not always in the right direction. All aspects of the hypotheses that concern the feature *number*, then, need further investigation.

The corpus investigation reported here provides a new contribution to the debate about what features count in intervention and what do not. As discussed in work by Franck and Villata, one approach defines the relevant notions of similarity as *narrow similarity*, where only morphosyntactic features count (Rizzi, 2004; Belletti et al., 2012). Another approach defines a notion of *broad similarity*, where any syntactic or semantic features can count, as long as they can be related to verb argument relations. (See, for example, Villata and Franck, which also show an effect of animacy in *wh*-islands). Our results seem to indicate that a more articulate characterisation of intervention locality is needed, as we find results compatible, but only partially, with both approaches. The distinction in strength of effect of the *number* feature between Italian and English and the effect of *type* feature is predicted by a narrow theory of similarity, that ties the effects and its strength to the morpho-syntactic make up of the language. The effect of *animacy*, though, extends the set of features relevant to intervention to lexical semantic aspects of the actants in grammatical long-distance dependencies.

These corpus results also join the rich current debate on the exact nature of structural dependencies and locality in computational method, and like other approaches, show for the moment, mixed conclusions. While some experiments have shown that Recursive Neural Networks can learn the main descriptive properties of long-distance dependencies in English, for example the fact that they obey a uniqueness constraint (only one gap per filler) and also that they obey island constraints (Wilcox et al., 2018), work attempting to replicate finer-grained human judgments for French have failed to show a correlation with human behaviour (Merlo and Ackermann, 2018), while other work on English has found mixed results (Chowdhury and Zamparelli, 2018). Lack of correlation with human grammaticality judgments has also been found in *wh*-islands and object relative clauses for both French and English (Merlo, 2019). More work will be needed to establish the exact boundaries of quantitative properties in long-distance dependencies across several languages.

6 Conclusions and future work

The contributions of this treebank study are many-fold. First, we formulate quantitative predictions about object-oriented relative clauses based on intervention theory. These predictions aim to identify which features come into play in defining the notion of intervener, and with what strength. Our results corroborate some previous findings concerning morphosyntactic features and animacy, but not all, opening the door to further investigation.

Future work will have to extend the investigation to other features and to other constructions that have been proposed and discussed in the theory and develop more complex models of intervention similarity. Current work is investigating the morpho-syntactic feature *person* and models of similarity related to word embeddings (Merlo, 2019).

Finally, thanks to the resources such as UD, we can also envisage to extend the investigation to the many languages for which theoretical predictions already exists and help formulate new ones in new languages.

Acknowledgments

The work was partially done while the first author was participating in the ERC Advanced Grant n. 340297 SynCart, which we gratefully acknowledge.

References

- Flavia Adani. 2012. Some notes on the acquisition of relative clauses: new data and open questions. In Valentina Bianchi and Cristiano Chesi, editors, *ENJOY LINGUISTICS! Papers offered to Luigi Rizzi on the occasion of his 60th birthday*, pages 6–13. CISCLPress.
- Lars Ahrenberg. 2015. Converting an English-Swedish parallel treebank to universal dependencies. In *Third International Conference on Dependency Linguistics (DepLing 2015)*, pages 10–19, Uppsala, Sweden, August. Association for Computational Linguistics.
- Adriana Belletti and Cristiano Chesi. 2014. A syntactic approach toward the interpretation of some distributional frequencies: Comparing relative clauses in Italian corpora and in elicited production. *Rivista di Grammatica Generativa*, 36:1–28.
- Adriana Belletti, Naama Friedmann, Dominique Brunato, and Luigi Rizzi. 2012. Does gender make a difference? Comparing the effect of gender on children’s comprehension of relative clauses in Hebrew and Italian. *Lingua*, 122(10):1053–1069.
- Anamaria Bentea. 2016. *Intervention effects in language acquisition: the comprehension of A-bar dependencies in French and Romanian*. Ph.D. thesis, University of Geneva.
- Ann Bies, Justin Mott, Colin Warner, and Seth Kulick. 2012. English web treebank. *Linguistic Data Consortium, Philadelphia, PA*.
- Cristina Bosco and Manuela Sanguinetti. 2014. Towards a Universal Stanford Dependencies parallel treebank. *CLARIN-D*, page 14.
- Cristina Bosco, Montemagni Simonetta, and Simi Maria. 2013. Converting Italian treebanks: Towards an Italian Stanford Dependency treebank. In *7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 61–69. The Association for Computational Linguistics.
- Silke Brandt, Evan Kidd, Elena Lieven, and Michael Tomasello. 2009. The discourse bases of relativization: An investigation of young German and English-speaking children’s comprehension of relative clauses. *Cognitive Linguistics*, 20(3):539–570.
- Shammur Absar Chowdhury and Roberto Zamparelli. 2018. RNN simulations of grammaticality judgments on long-distance dependencies. In *Proceedings of the 27th International Conference on Computational Linguistics (COLING’18)*, pages 133–144. Association for Computational Linguistics.
- Ulrich Hans Frauenfelder, Juan Segui, and Jacques Mehler. 1980. Monitoring around the relative clause. *Journal of Verbal Learning and Verbal Behavior*, 19(3):328–337.
- Naama Friedmann and Rama Novogrodsky. 2004. The acquisition of relative clause comprehension in Hebrew: A study of SLI and normal development. *Journal of Child language*, 31(3):661–681.
- Naama Friedmann, Adriana Belletti, and Luigi Rizzi. 2009. Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua*, 119(1):67 – 88.
- Edward Gibson. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68:1–76.
- Nino Grillo. 2008. *Generalized minimality: Syntactic underspecification in Broca’s aphasia*. Ph.D. thesis, University of Utrecht.
- Katri Haverinen, Jenna Nyblom, Timo Viljanen, Veronika Laippala, Samuel Kohonen, Anna Missilä, Stina Ojala, Tapio Salakoski, and Filip Ginter. 2013. Building the essential resources for Finnish: the Turku dependency treebank. *Language Resources and Evaluation*, 48(3):493–531.
- Paola Merlo and Francesco Ackermann. 2018. Vectorial semantic spaces do not encode human judgments of intervention similarity. In *Proceedings of the 22nd Conference on Computational Natural Language Learning, CoNLL 2018*, pages 392–401, Brussels, Belgium, October.

- Paola Merlo. 2016. Quantitative computational syntax: some initial results. *Italian Journal of Computational Linguistics*, 2(1):11–30.
- Paola Merlo. 2019. Probing word and sentence embeddings for long-distance dependencies effects in French and English. In *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, Florence, Italy, July. Association for Computational Linguistics.
- Joakim Nivre. 2015. Towards a universal grammar for natural language processing. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 3–16. Springer.
- Jean-Yves Pollock. 1989. Verb movement, universal grammar, and the structure of IP. *Linguistic Inquiry*, 20:365–424.
- Luigi Rizzi. 1990. *Relativized Minimality*. MIT Press, Cambridge, MA.
- Luigi Rizzi. 2004. Locality and left periphery. In Adriana Belletti, editor, *The cartography of syntactic structures*, number 3 in Structures and beyond, pages 223–251. Oxford University Press, New York.
- Luigi Rizzi. 2013. Locality. *Lingua*, 130(1):69 – 86.
- Douglas Roland, Frederic Dick, and Jeffrey L Elman. 2007. Frequency of basic English grammatical structures: A corpus analysis. *Journal of memory and language*, 57(3):348–379.
- Sandra Villata and Julie Franck. 2016. Semantic similarity effects on weak islands acceptability. In *41st Incontro di Grammatica Generativa Conference*, Perugia, Italy. <https://archive-ouverte.unige.ch/unige:82418>.
- Ethan Wilcox, Roger Levy, Takashi Morita, and Richard Futrell. 2018. What do RNN language models learn about filler–gap dependencies? In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 211–221. Association for Computational Linguistics.