IJCoL

Italian Journal of Computational Linguistics

Rivista Italiana di Linguistica Computazionale

Volume 2, Number 1
june 2016

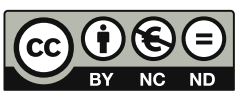
Emerging Topics at the Second Italian Conference on Computational Linguistics

aAccademia university press

Registrazione presso il Tribunale di Trento n. 14/16 del 6 luglio 2016

Rivista Semestrale dell'Associazione Italiana di Linguistica Computazionale (AILC)

direttore responsabile
Michele Arnese



isbn 978-88-99200-99-2

Accademia University Press
via Carlo Alberto 55
I-10123 Torino
info@aAccademia.it
www.aAccademia.it/IJCoL_02

aAccademia university press

IJCoL

Volume 2, Number 1
june 2016

Emerging Topics at the Second Italian Conference on Computational Linguistics

CONTENTS

Quantitative computational syntax: some initial results

Paola Merlo*
University of Geneva, Switzerland

In the computational study of human intelligence, the language sciences are in the unique position of resting both on sophisticated theories and representations and on large amounts of observational data available for many languages. In this paper, we discuss some recent results, where large-scale, data-intensive computational modelling techniques are used to address fundamental linguistic questions on the quantitative properties of abstract grammatical representations. Specifically, we present a programme of research exemplified in three case studies to identify the causes of frequency differentials. In the area of word order, we discuss work that investigates whether typological and corpus frequencies are systematically correlated to abstract syntactic structures and to higher-level structural principles of minimisation and efficiency. In the area of verb meaning, corpus-based computational models are discussed that investigate how frequencies are correlated to well-known lexical effects in causative alternations and morphological marking. The large corpus-based, cross-linguistic component of the work and the abstract grammatical hypotheses on word order and verb meaning provide new empirical and computational evidence to the important debate on language variation, its extent and its limits and illustrate how to bring corpus-based computational methodology to bear on theoretical syntactic issues. In so doing, we help reduce the current gap between theoretical and computational linguistics.

1. Charting the Frequency Landscape

Quantitative methods and corpus-based data collection have been used extensively in the study of language acquisition, language processing, and historical linguistics, and they figure prominently in usage-based linguistic explanations (Bybee 2007), but they have played very little role in the reasoning and theorising about the representations and the rules of formal grammars, with the exception of the notion of markedness (Moravcsik and Wirth 1983; Haspelmath 2006)[1]. Current syntactically annotated cor-

* Department of Linguistics - Rue de Candolle 5, 1204 Geneva, Switzerland.
E-mail: Paola.Merlo@unige.ch

1 A paper like this draws on two very different scientific traditions, merging formal linguistic and computational concepts. Rather than slowing down the flow of argumentation in the text, I will provide some lengthy footnotes to clarify some concepts, as needed. The notion of markedness will be mentioned several times, so I start with it. Haspelmath (2006) identifies twelve different uses of the word *markedness*, so only a vaguer definition will be given here. Initially, the notion of marked/unmarked elements was introduced in phonology to indicate the element of a contrastive pair that was phonologically or morphologically marked. For examples, in most European languages, the singular is unmarked and the plural is explicitly marked by a plural ending. Markedness refers also to the element of a contrastive pair that is more difficult to process or learn, or that is less frequent, in a language or across languages. It has been shown that these different notions are distinct but correlated: a marked element is usually less frequent, less productive, and less attested typologically. I refer to Haspelmath (2006) for a very systematic and interesting discussion.

pora allow us to ask foundational questions on the role of frequency and quantitative data in the theories of grammar[2]. It has been shown, for example, that subcategorisation frequencies in corpora are correlated to grammaticality judgements (Merlo 1994), that deep principles of verbal lexicon organisation, namely verb classes, show robust statistical regularities within and across languages, and that this is because these frequencies are surface reflexes of underlying thematic regularities (Merlo and Stevenson 2001; Merlo et al. 2002). These same frequency-based features have been shown to be at work in the learning of verb meaning in children (Scott and Fisher 2009; Clegg and Shepherd 2007). These results have been followed by many other investigations in the computational linguistics community in verb classification, sense disambiguation, propositional annotation, distributional lexical semantics, among other topics (Palmer, Gildea, and Kingsbury 2005; Schulte im Walde 2006; Abend, Reichart, and Rappoport 2008; Baroni and Lenci 2010), that also show a close connection between quantitative surface properties and abstract syntactic and semantic representations.

Differences in linguistic frequencies become then the fact that needs explanation. Frequency is a puzzling property of language constructs, whose correlation with other aspects of grammatical representations or other linguistic observations is not clear. Functionalists and formal grammarians are in agreement in assuming that frequencies are an expression of language use. They then treat the relationship between grammar and frequency in a very different way. Functionalist approaches have addressed the relationship between frequency and grammar by assuming that usage shapes grammar, and that frequency of use is the cause of some prominent linguistic effects, especially related to change. (See (Bybee 2007), among many others.) According to Haspelmath (2006), "Frequency of use is a property of parole or performance, not of language structure or competence, and throughout the 20th century most linguists have shown little interest in explaining structure in terms of use." (page 16).

From a generative or cognitive point of view, frequencies are not part of the grammar or the cognitive system. A very well known point of view denies its relevance to any linguistic investigation (Chomsky 1965). This point of view assumes that frequency-based, quantitative properties of text are totally unrelated to the underlying grammatical representations of language that linguistic theory proposes. This standpoint makes the clear prediction that data-driven, statistical methods would not work well at all in producing grammatical representations automatically. This is clearly incorrect. In light of the more recent linguistic results briefly mentioned above and especially in light of the obvious and pervasive success of statistical methods in data-driven natural language processing, we will discard this standpoint as too reductionist.

Another point of view relates frequency to the notion of markedness, both in the generative and the functional sense. Markedness has been equated in generative grammar to the cost of structure building or movement operations (Cinque 2005, 2013) or to default parameter setting (Travis 1984; Baker 2002; Yang 2003). The functionalist notion of markedness — markedness as complexity, markedness as difficulty, markedness as abnormality, markedness as multidimensional correlation —is related to frequency as the zero-coded, most flexible, typologically flexible element (Haspelmath 2006). Frequency distributions and their direct relation with structural recursion or their inverse relation with some notions of acquisition or processing complexity have also been ex-

2 By frequency, we mean in this paper *counts* in a corpus, the counts of a given construction, for example, or the counts of a property in typological data, for example, the number of languages that exhibit a certain word order.

plained as an effect of pressure for efficient communication (Dryer 1992, 2009; Hawkins 1994, 2004; Gibson 1998; Tily, Frank, and Jaeger 2011; Fedzechkina, Jaeger, and Newport 2012; Zipf 1949).

Current large-scale, syntactically-annotated resources for several languages allow us to extend these investigations to the correlation between quantitative linguistic properties and abstract linguistic representations and operations. I want to put forward here a point of view on frequency that differs both from the functionalist approach, where frequency and usage shape grammar — an approach where frequency is the cause and linguistic phenomena are the effect — and also differs from the traditional formal grammar disregard for quantitative aspects of formal representations. Frequency is neither the independent variable in the explanation nor irrelevant to language. I will here illustrate results that speak to the claim that *frequencies are grammatical facts that require explanation*: they are the *dependent* variable in a model and they are the expression of a grammar that comprises a quantitative component, be it weights on its operations and combinatorics or probabilities of its properties. I illustrate this approach by grounding explanations about major syntactic properties, such as the order of words in the sentence or the lexical semantic primitives of verbs, in observational empirical evidence. The specific quantitative property we will study is frequency of occurrence, cross-linguistically as found in typological surveys and within a language, as found in text.

2. Three case studies on frequency

Our research objectives leverage both frequencies across languages (typological distributions) and quantitative properties within languages (based on large corpora, for many languages). We hypothesize the same root causes for both these types of frequencies, thereby predicting specific correlations between them. We discuss results that *predict* frequencies of two linguistic phenomena: syntactic word order and the causative alternations, within and across languages. We ask for example, why do adjectives occur more often postnominally than prenominally across the world languages? Why is it that Romance languages prefer to position adjectives prenominally if they are followed by a prepositional phrase? When do languages prefer to use a morphologically complex form in transitive constructions and when do they prefer a morphologically complex intransitive construction?

The work reported here illustrates two possible explanatory assumptions for frequency distributions: weighted complexity-based explanations and probability-based explanations.

On the one hand, if frequencies are the linguistic object of study — they are the dependent variable of the model — the independent variable could be some form of complexity. Complexity can in turn be defined in two ways: complexity of operations and complexity of representations. Two of the works we discuss investigate the link between frequencies of syntactic word orders and underlying grammatical principles and structures. One of the works we discuss here spells out an approach where the frequency is generated by the complexity of the operation. Specifically, the complexity of basic syntactic operations explains the frequency with which different orders of the elements of a noun phrase occurs across languages of the world, typologically. Another piece of work elaborates an approach where frequency of alternative word orders in a given language are determined by the complexity of the representations. Specifically, the complexity of the tree representations, and even more precisely the length of the dependencies in the tree, determine the preferred orders of adjectives in the noun phrase.

In short, the hypothesis is that the differential costs of basic syntactic operations and different underlying structures yield differential frequencies in typological distributions of word order and in the distribution of word order variants within a language (Merlo 2015; Gulordava, Merlo, and Crabbé 2015; Gulordava and Merlo 2015b), and (Merlo and Ouwayda, forthcoming).

On the other hand, we also discuss a body of work that investigates the link between frequencies and components of verb meaning and demonstrates the hypothesis that probabilistic lexical semantic primitives in causative alternations yield differential surface structural frequencies (Samardžić and Merlo 2010; Samardžić and Merlo 2012; Samardžić 2013) and Samardžić and Merlo (to appear). This line of work explores then a different kind of explanation for the frequency of linguistic phenomena: frequencies are the observed effect of an underlying inherently probabilistic system.

Notice that these two kinds of explanations are not coextensive: a complexity-based explanation could very well be quantified without any notion of probabilty (as in the second case study) and hence any recourse to a notion of uncertainty in the system.

2.1 Statistical Word Order Universals as Complexity of Operations

One of the most easily observable distinguishing features of human languages is the order of words: the position of the verb in the sentence or the respective order of the modifiers of a noun, for example. For each of these linguistic phenomena, word orders vary greatly cross-linguistically, but each language has very strong preferences for a few or only one order, and, across languages, not all orders are equally preferred (Greenberg 1966; Dryer 1992; Cinque 2013; Baker 2002). Very many theories and descriptions have attempted to explain word order differences and similarities (called word order universals). Current explanations of typological word order universals have moved beyond trying to define possible or impossible word orders and are aiming at finer-grained distinctions to predict the whole frequency distributions of attested word orders (Cysouw 2010).

One such area is the study of the underlying complexity that gives rise to different word orders across languages in the internal structure of Noun Phrases, and the observational universal called Universal 20 (Greenberg 1966) – the universal governing the linear order of a noun and its modifiers[3].

> **Greenberg's Universal 20**
> When any or all the items (demonstrative, numeral, and descriptive adjective) precede the noun, they are always found in this order. If they follow, the order is exactly the same or its exact opposite.

Greenberg's formulation identifies four main elements in the noun phrase: N, Dem, Num, Adj. Universal 20 states, explicitly or implicitly, many typological properties of NPs. First, it states that, of the 24 logically possible orders, only three are attested. Second, it states that two postnominal orders but only one prenominal order are attested, thereby establishing an asymmetry between prenominal and postnominal distributions of word orders. Third, it states that of the three possible orders, the prenominal order is the mirror image of one of the postnominal orders.

3 This line of work started in Merlo (2015) , summarised here, and currently is being developed in Merlo and Ouwayda (forthcoming) to finer-grained theoretical detail.

Table 1
Attested word orders of Universal 20 and their counts: the first two columns report counts of genera and languages in Dryer's sample, the last two columns report counts from Cinque (2005) and from a large sample (Cinque, 2013, p.c.). (See text for more explanation.)

				Dryer's Languages	Dryer's Genera	Cinque's 05 Languages	Cinque's 13 Languages
Dem	Num	Adj	N	74	44	V. many	300
Dem	Adj	Num	N	3	2	0	0
Num	Dem	Adj	N	0	0	0	0
Num	Adj	Dem	N	0	0	0	0
Adj	Dem	Num	N	0	0	0	0
Adj	Num	Dem	N	0	0	0	0
Dem	Num	N	Adj	22	17	Many	114
Dem	Adj	N	Num	11	6	V. few (7)	35
Num	Dem	N	Adj	0	0	0	0
Num	Adj	N	Dem	4	3	V. few (8)	40
Adj	Dem	N	Num	0	0	0	0
Adj	Num	N	Dem	0	0	0	0
Dem	N	Adj	Num	28	22	Many	125
Dem	N	Num	Adj	3	3	V. few (4)	37
Num	N	Dem	Adj	5	3	0	0
Num	N	Adj	Dem	38	21	Few (2)	180
Adj	N	Dem	Num	4	2	V. few (3)	14
Adj	N	Num	Dem	2	1	V. few	15
N	Dem	Num	Adj	4	3	Few (8)	48
N	Dem	Adj	Num	6	4	V. few (3)	24
N	Num	Dem	Adj	1	1	0	0
N	Num	Adj	Dem	9	7	Few (7)	35
N	Adj	Dem	Num	19	11	Few (8)	69
N	Adj	Num	Dem	108	57	V. many (27)	411

The frequencies of occurrence of the 24 combinatorially possible orders of the four elements on the Noun Phrase have been discussed in several publications (Dryer 2006, Cinque 2005, 2013) and are shown in Table 1: the first two columns are Dryer (2006)'s counts by language and by genera; and the following columns are Cinque's counts, as can be deduced from the 2005 paper and more recent 2013 counts (personal communication)[4]. Despite differences across the different counting methods and across authors, the rank of languages or genera based on frequencies is almost identical, and there is therefore substantial agreement on which word orders are frequent or very frequent, and which are rare or unattested.

Some aspects of Greenberg's formulation have been confirmed by the collection of these larger samples (Cinque's and Dryer's samples comprise hundreds of languages),

4 In the models that we present later, we use Cinque's most recent counts. We report here Dryer's counts by languages and by genera to illustrate the fact that taking into account density of linguistic neighbourhood in the sample (i.e. counting genera rather than single languages) does not fundamentally change the typological distribution of counts. In Dryer's definition "a genus is a group of languages whose relatedness is fairly obvious without systematic comparative analysis and which even the most conservative "splitter" would accept." Examples are such subfamilies of Indo-European as Germanic, Slavic, and Romance languages.

but some others have been found to be too strong. For example, the larger, more recent samples confirm that there are more postnominal orders than prenominal orders and that the two mirror-image orders are the most frequent. On the other hand, many more orders are attested than the three mentioned in Universal 20 and the third order mentioned in Universal 20 (N Dem Num Adj) is not as frequent as Greenberg predicted. Most importantly, the observation that not all possible orders are equally frequent is an observation that requires explanation.

Several authors have attempted to reconcile generative mechanisms with typological observations by a system of costs and constraints that generate statistical universals (Cinque 2005; Abels and Neeleman 2009; Cinque 2013; Steedman 2011; Steddy and Samek-Lodovici 2011). Perhaps the best known paper of this kind is Cinque's (Cinque 2005), where Greenberg's Universal 20 is derived from independently motivated principles of syntax organised in a derivational explanation. Cinque proposes that the actually attested orders, and none of the unattested ones, are derivable from a single universal order (the order Dem Num Adj Noun), and from independent conditions on phrasal movement. Different types of movement can move the elements to different positions in the phrase: all the way to the beginning of the phrase or only partially. Different forms of movement are more costly than others and 'no movement' is the preferred unmarked option. By a carefully calibrated assignment of costs to the movement operations, Cinque's proposal also derives the exceptions, and the different degrees of markedness of the various orders.

In a different proposal, a factorial, but not derivational, explanation is proposed (Cysouw 2010). An explanation of typological frequencies is produced by the cumulative combination of a statistical model. Three characteristics are used by the models: hierarchical structure, noun-adjective order, and whether the noun is at the phrase boundary. This factorial explanation does not provide a generative process that explains how the different word orders could arise from a common grammar, but it identifies observable predictive properties of the frequency distributions of word order and their relative importance.

Dryer proposes a factorial explanation based on general principles of symmetry and harmony (Dryer 2006, 2009). Differently from Cinque's and Cysouw's this proposal does not assign weights to the factors. The factors comprise two symmetry principles that describe the closeness of the modifiers to the noun: Symmetry Principle 1 describes the preference of the adjective and numeral to occur closer to the noun than the Determiner, and symmetry principle 2 states that the adjective tends to occur closer to the Noun than the Numeral, when they occur on the same side of the Noun. Dryer also uses a principle of asymmetry that captures the main observation that prenominal modifiers exhibit fewer alternatives than post-nominal modifiers (also observed by Cinque); a principle of intra-categorial harmony, which says that all modifiers tends to occur on the same side of the noun; and Greenberg's universal 18[5].

These three models offer different views of grammatical complexity and how it can be related to frequency. Cinque's proposal is a traditional derivational theory of complexity (DTC) applied to language typology. In psycholinguistics, the derivational theory of complexity was a linking hypothesis stating that difficulty in the perception of a sentence is related to how many derivational operations must be applied to derive the surface structure, according to the rules of transformational (derivational) grammar

5 Greenberg's Universal 18: When the descriptive adjective precedes the noun, the demonstrative and the numeral, with overwhelmingly more than chance frequency do likewise.

(Miller and McKean 1964). Cinque proposes an augmented DTC-theory where different types of operations have differential costs, to explain frequency distributions: the global complexity of an order is an additive function of the number of movement operations weighted by their costs. The reasoning behind the linking between markedness and frequency is that more costly word orders will be dispreferred in production, comprehensions and possibly acquisition, and hence be less frequent. In this respect, Cinque joins Hawkins's proposals (Hawkins 1994, 2004) in connecting processing loads to typological distributions, but it differs from Hawkins's proposal, as we will see below, in the underlying notion of complexity. Another important property of Cinque's proposal is that it assumes that the underlying movement operations that give rise to the typological distributions are not directly observable. The model is therefore only partly observable, it is a *hidden* model. The distinction between hidden and observable models is important. A hidden model assumes that an important part of the structure of the model can be learned, or at least its parameters can be estimated, without direct observation. A hidden model is more appropriate to represent theories with general abstract principles.

Differently from Cinque, Cysouw proposes that word order depends on a small number of observed independent variables: whether adjective and noun are adjacent or whether the noun is at the edge of the phrase. Unlike Cinque's and Dryer's, Cysouw's model has no underlying variables (technically this is called a fully observed model). Complexity is equated to the weights of the observed independent variables.

Dryer proposes a set of explanatory properties —properties such as symmetry and harmony— whose characteristic is that they concern the linguistic system as a whole. Violations of these properties are then added up to predict the distribution in frequency. This model is observed, in that it uses general properties derived from surface word orders, but the properties apply to the whole set of the world's languages, and cannot be directly determined by inspection of only a single language or a single word order.

To compare the explanatory adequacy of these models, Merlo (2015) defines a formal encoding that maps the derivational operations in Cinque's proposal and Cysouw's and Dryer's factors to a feature-vector representation. Then, a model is learnt through a standard learning algorithm on a subset of the data; finally, the model is run on previously unseen data to test the generalisation ability of the constructed models. In short, each word order is represented by a "summary" feature vector, which encodes the theory. These vectors are then used by a supervised classifier to predict the word order frequency of unseen word orders.

We apply this same idea, using simple linear regression, to compare these different notions of complexity as predictors of frequency counts: as cost of movement (Cinque 2005, 2013); as weights of observable constraints on adjacency and positioning of the elements in the Noun Phrase (Cysouw 2010); as violations of global notions of symmetry and harmony that apply to the whole system (Dryer 2006). Taking the generalising ability of the proposals on unseen data as our measure of explanatoriness, we can compare these different approaches to complexity. We run linear regressions for these three models and look at their correlations and the properties that are used to predict the orders. Cinque's model is shown in (1), Cysouw's in (2) and Dryer's in (3). We refer to Cinque's original paper for a detailed explanation of the different kinds of movement. What is relevant here is that each factor is a different kind of movement, and that not all kinds of movements are used in the model. The same is true for Cysouw's and Dryer's models, where one the factors is not used. Clearly, Cinque's and Dryer's model have better fit to the data, as shown by the correlation coefficients.

(1) Correlation coefficient 0.64

Frequency= -105.12 x NP movement without pied-piping +
-127.32 x NP movement, pied-piping [XP [NP]] +
-99.70 x PartialMove +
-111.39 x NP-splitting movement +
-148.31 x Moves a phrase not containing NP +
+ 234.84

(2) Correlation coefficient 0.23

Frequency= 62.46 x NA-adjacency +
88.23 x Dem-edge +
-2.95

(3) Correlation coefficient 0.59

Frequency= 130.30 x symmetry1 +
107.94 x symmetry2 +
106.66 x IC-harmony +
-130.15

The results are interesting. Cysouw's limited ability to generalise might indicate that principles developed specifically to explain Universal 20 are not good expressions of the general notion of complexity. For example, many languages in the world exhibit an SVO sentential order, where the main element, the verb, is medial. It appears, then, that a general preference for being at the edge of the phrase is not likely (and it is also in contradiction with Dependency Length Minimisation effects, see below). Dryer's and Cinque's proposals show better correlation with typological frequencies, a result also confirmed by classification (Merlo, 2015). But for example, an error analysis of Cinque's results shows that complete movement and even no movement at all are more costly than partial movement. Not only is this different from Cinque's assumption and his hand-assigned weights, but it also indicates that a "derivational theory of complexity" account is not supported. A DTC account would predict that no movement is easier than partial movement which is easier than complete movement, directly proportional to the derivation operations needed. A more articulated linking hypothesis between derivational operations and frequency is needed. Some syntactic proposals exist where the cost of movement is specific to the Noun Phrase and depends on the reasons to move. For example, if movement is needed to trigger agreement then even partial movement might not be costly (Shlonsky 2012). Preliminary results by the author's group inspired by these theories indicate that models of complexity need to be developed at a finer-grain of detail, with complexities specific to the moved element and the constructions in which they are moved (Merlo and Ouwayda, forthcoming)[6].

6 Interested readers might want to know how interactions between the factors of the regression model affect the results. This specific issue is tested and discussed in detail in Merlo (2015) and Merlo and Ouwayda (forthcoming). Basically, the conclusion is that the features in these models can, by and large, be considered independent.

Assuming that the most frequent word orders have lowest cost of derivation or lowest deviance from some general principle of grammatical organisation is strongly correlated to a view where frequency is an expression of markedness, like Haspelmath's proposes (Haspelmath 2006). The fact, however, that there is no perfect correlation between the underlying principle determining "markedness" and frequency indicates that, contra Haspelmath, the two cannot be completely equated. This opens the door to investigations of other causes of frequency differentials.

2.2 Frequency of Alternative Word Orders as Structural Locality

Typological theories of word order have concentrated on predicting dominant word order. A word order is considered dominant if it is the most frequent, or if it is the preferred order in main assertive, neutral clauses or is otherwise demonstrably the 'default' basic word order[7]. But very important information is lost about how dominant the order really is, since we don't know the distribution of possible word order alternatives in a given language. By looking at collections of annotated corpora, we can work with complex frequency distributions without the simplifying concept of a dominant word order.

If we want to study the distributions of alternative word orders in a language, we need to study the factors that might determine the preference for one or the other order. In looking at the distributions of word orders in a corpus, a quantitative tendency has been observed across many languages and constructions: if two alternative orders are possible, the one which puts the short elements first is preferred, as illustrated in the examples below (Bresnan et al. 2007; Wasow 2002), where the b option are preferred.

(4) Alternation between NP and PPs
  a. He put [on the table] [the groceries bought at the market]
  b. He put [the groceries bought at the market] [on the table]

(5) Alternation between PPs
  a. He talked [to Mary] [about last week's exam]
  b. He talked [about last week's exam] [to Mary]

(6) Verb-particle construction
  a. I threw [the bag] out
  b. I threw out [the bag full of garbage]

Through a series of corpus analyses of prepositional phrases and relative clauses, among other constructions, (Hawkins 1994, 2004) shows that syntactic choices generally respect the preference for placing short elements closer to the head than long elements. Hawkins's work is representative of much work on language processing which attributes parsing performance to the distance or locality of linguistic constituents and their dependents (Gibson 1998, 2000; Lohse, Hawkins, and Wasow 2004; Demberg and Keller 2008).

If we analyse the structure of the sentences, we actually notice that a more general principle, called Dependency Length Minimisation (DLM) is at work (Temperley 2007). The DLM principle states that if there exist possible alternative orderings of a phrase, the

7 The World Atlas of Language Structure uses a notion of word order based on frequency. See the detailed explanation in Dryer (2013)

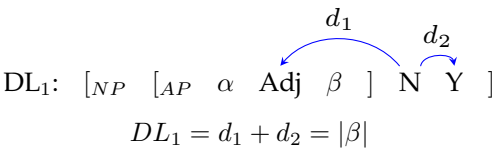


Figure 1
Prenominal adjective

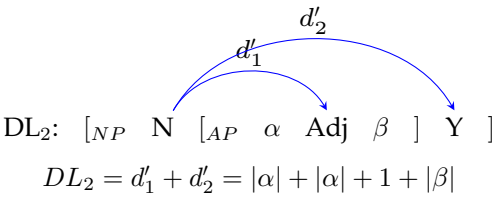


Figure 2
Postnominal adjective

one with the shortest overall dependency length (DL) is preferred, where the length of a dependency is measured as the number of words between the head and its dependent. Consider, for example, the case illustrated by the first example above, when a verb has both a direct object (NP) and a prepositional complement or adjunct (PP). Two alternative orders of the verb complements are possible: VP_1 = V NP PP, whose length is DL_1 and VP_2 = V PP NP, whose length is DL_2. If DL_1 is bigger than DL_2, then VP_2 is preferred over VP_1.

We currently have many large-scale, syntactically hand-annotated treebanks. These data allow us to verify these claims within a language and for many languages. Recently, global measures of dependency length on a larger scale have been proposed, and cross-linguistic work has used these measures and demonstrated their minimisation (Gildea and Temperley 2010; Futrell, Mahowald, and Gibson 2015; Gulordava and Merlo 2015a).

DLM has often been explained functionally, as a processing effect that aims to reduce memory or processing load. It is not clear what its impact would be in the structurally more limited nominal domain: it might appear that the short distances of elements inside the noun phrase do not lend themselves to exhibiting DLM effects. Concentrating on Romance languages, (Gulordava, Merlo, and Crabbé 2015; Gulordava and Merlo 2015b) have recently investigated short spans, studying DLM effects in the noun phrase (the domain of Universal 20 and Universal 18). We discuss here the results of their studies[8].

Specifically, choosing languages that allow enough pre/post-nominal adjective placement variation, Gulordava and colleagues look at the structural factors that play a role in adjective-noun word order alternations in Romance languages. They assume a simplified noun phrase with only one adjective modifier adjacent to the noun and two possible placements for an adjective phrase: post-nominal and prenominal. The adjective modifier can be a complex phrase with both left and right dependents (α and

8 They use the dependency annotated corpora of five Romance languages, from the multilingual CoNNL 2006 and 2009 shared tasks or from more recent universal dependencies annotations: Catalan, Spanish, Italian (Hajič et al. 2009), French (McDonald et al. 2013), and Portuguese (Buchholz and Marsi 2006).

β, respectively). The noun phrase can have a right modifier Y. The structures for the possible cases are shown in Figures 1 and 2. These structures correspond to examples like those shown in (7), in Italian (Adj='bella', N='casa', Y= 'al mare')[9].

(7) a. ... vede la bella casa al mare. ('... sees the beautiful house at the sea')

b. ... vede la casa bella al mare. ('... sees the house beautiful at the sea')

Gulordava and colleagues develop a mixed-effects model to test which of the fine-grained predictions derived from DLM are confirmed.[10] The different elements in the DLM configuration are encoded as three factors: corresponding to the elements illustrated in Figures 1 and 2 and example (7): *LeftAP* - the cumulative length (in words) of all left dependents of the adjective, indicated as α in Figures 1 and 2; *RightAP* - the cumulative length (in words) of all right dependents of the adjective, indicated as β in Figures 1 and 2; *RightNP* - the indicator variable representing the presence or absence of the right dependent of the noun, indicated as Y in Figures 1 and 2.

Their findings mostly confirm the DLM predictions about adjective placement with respect to the noun given the adjective dependents: presence of preadjectival material α triggers a preference for prenominal adjectives while presence of postadjectival material β yiels a preference for postnominal adjectives. They also demonstrate that it is not only the length of the adjective-noun dependency that is at stake, but also the interaction with the surrounding dependencies. More precisely, it is demonstrated that the presence of a right dependent of the noun, Y in the pictures, affects the position of the adjective which modifies this noun: the prenominal position is more often preferred in such cases.

These results confirm that DLM is active in noun phrases, despite the shortness of the dependencies, often only one word long. It is open to debate, however, whether DLM is supported as a general principle of cognition, or is undermined, given that such short dependencies weaken the plausibility of the memory-based functional explanation that has been offered for it.

In terms of approaches to complexity, DLM is a structural locality theory of complexity, similar in spirit, if not in detail, to the Relativised Minimality restrictions to structures — which requires long-distance relations to be minimal and not to be interrupted by an intervening element —or to the dependency locality theory of sentence processing— the complexity of integrating a word in an existing structure depends on

9 Gulordava and Merlo (2015) also consider the direction of dependency of the NP itself, that is they also contrast the examples in (7) to examples such as *La bella casa al mare è vuota* ('the beautiful house at the sea is empty') *La casa bella al mare è vuota*. ('the house beautiful at the sea is empty'). Their calculations of DLM predictions is also more complex and distinguishes four cases. They found, however, that the position of the governor of the NP does not yield significant interactions in their model, so we do not discuss it here.

10 A mixed-effect model is a statistical model containing both fixed effects and random effects. Mixed-effect models are applied in many disciplines where multiple correlated measurements are made on each unit of interest. More precisely, Generalized Linear Mixed Models describe an outcome as the linear combination of fixed effects X and conditional random effects Z associated with grouping of instances, where β and γ are the corresponding weights of the effects.

$$y = X\beta + Z\gamma + \epsilon \tag{2}$$

In logistic regression models, this linear combination is then transformed with the logit link function to predict the binomial output:

$$Order = \frac{1}{1 + \exp^{-y}} \tag{3}$$

In our model, $Order = 0$ codes the prenominal adjective order and $Order = 1$ codes the postnominal order.

the length of the dependency that is being built to integrate the word (the distance between the two elements being integrated) (Rizzi 1990; Gibson 1998). These approaches assume that longer dependencies are harder, or even ungrammatical. Their view of complexity is not derivational. Two identical structures have the same complexity, irrespective of the number of operations used to build it. In this respect, they contrast with those views where all the complexity resides in the derivation and most of the grammatical distinctions are encoded by movement operations. If taken as a theory of complexity, DLM makes a prediction about word order frequencies in a language, for those constructions that allow alternations. DLM states that longer dependencies, overall, are dispreferred. Consequently, minimised (sub)structures should be more frequent than less minimised alternatives, ceteris paribus.

The work described so far investigates quantitative factors underlying typological and corpus variation of word order, to discover the operations and principles that govern this variation. The two investigations study typological and corpus data separately. The phenomenon of causative alternations provides interesting evidence for the interactions between within-language corpus frequencies and cross-linguistic distributions of morphological markings, as discussed in the next section.

2.3 Frequency of verb alternations as lexical semantic probability

An interesting area of interaction between typological distributions and corpus frequencies has recently been found in the causative alternation (Samardžić and Merlo 2010; Samardžić and Merlo 2012; Samardžić 2013; Haspelmath et al. 2014; Heidinger 2015). We discuss here the results by Samardžić and colleagues.

The *causative(/inchoative) alternation* has been recognised in the linguistic literature as a wide-spread linguistic phenomenon, attested in almost all languages (Schaefer 2009). This alternation involves verbs such as *break* in (8), which can be realised in a sentence both as transitive (8a) and as intransitive (8b).

(8) a. *Causative*: Adam broke the vase.
b. *Anticausative*: The vase broke.

The causative alternation is wide-spread and frequent, but it appears in a large variety of lexical and morphological realisations across languages, whose underlying regularities are hard to define. Lexically, the sets of verbs that alternate are different across languages: while most of the alternating verbs are lexical counterparts across many languages, there are still many verbs which alternate only in some languages. Examples of only anticausative and only causative verbs in English are given in the top panel of Table 2, taken from (Alexiadou 2010). As the examples show, the English verbs *arrive* and *appear* do not alternate: their transitive realisation (*causative* in Table 2) is not available in English. However, their counterparts in Japanese, or Salish languages, for example, do alternate and are found both as transitive and intransitive. Similarly, the verbs such as *kill* are only found as transitive in English, while their counterparts in Greek or Hindi, for example, can alternate between intransitive and transitive use.

Often lexical causative are morphologically marked. But languages also differ in the morphological realisation of the alternation. Some examples of morphological variation are given in the bottom panel of Table 2, taken from (Haspelmath 1993). The morpheme that marks the alternation can be found on the intransitive form (Russian in Table 2), or the transitive form, like Mongolian. There are also languages where both forms receive

Table 2
Availability of the alternation (Alexiadou 2010) and morphological marking (Haspelmath 1993) in some examples of verbs and languages.

Availability:

	Causative	Anticausative
arrive, appear	+Japanese, +Salish, **-English**	+all languages
kill	+all languages	+Greek, +Hindi, **-English**

Morphological marking:

	Causative	Anticausative
Mongolian	xajl-**uul**-ax 'melt'	xajl-ax 'melt'
Russian	rasplavit' 'melt'	rasplavit'-**sja** 'melt'
Japanese	atum**-eru** 'gather'	atum**-aru** 'gather'

an (anti)causative marker, like in Japanese. English, on the other hand, is an example of a language where the alternation is not morphologically marked.

Many theories of the causative alternation explain this construction and its properties by postulating that verbs have sublexical features in their lexical entry (e.g. Reinhart's +c feature) that give rise to the causative and anticausative use expressed as a transitive or intransitive construction, respectively.

Given the variability of the sets of alternating verbs and the morphological markings illustrated above, questions that arise are: (i) What components of meaning do we find in the meaning of verbs that exhibit the causative alternation? (ii) What components of meaning regulate the morphological marking? (iii) What are the quantitative properties of these components?

The kind of answer Samardžić and colleagues pursue draws on previous results that tie quantitative patterns of syntactic variation to the underlying grammatical properties (Merlo and Stevenson 2001, ; Bresnan, Dingare, and Manning 2001; Bresnan 2007)). This previous work has established that verb classes (Levin 1993) show statistical regularities that are strong enough to be predictive and that are observable through carefully chosen and theoretically justified textual correlates, such as animacy, transitive use and strength of causative use. Applying this correlation inductively, one can conclude that differences in frequencies of alternations within and across languages are the expression of underlying components of meaning.

In a collection of papers on the use of corpus-based evidence for theoretical linguistic research and in Tanja Samardžić's dissertation, Samardžić and Merlo propose that we can conceptualize hidden lexical features (in the sense of hidden explained above)

expressing components of meaning of a verbal root as a probability or a gradient score. In the specific case of the causative alternation, inspired by Reinhart's +/-c feature (Reinhart 2002), they call this underlying property "likelihood of external causation". They argue that this feature determines whether a verb will preferentially surface as transitive or intransitive and also which of the two forms will be morphologically marked. If the likelihood of external causation is high, then the causative alternants will be more frequent and the anticausative form will be the marked form of the pair. Morphologically, the markedness of the anticausative will be expressed by the fact that a sample of verb equivalents across languages will show morphological marking of the anticausative in a majority of these languages. If the likelihood of external causation is low, the converse is observed. This is demonstrated by showing that the ratio of transitive to intransitive uses of the causative construction in a corpus of English mirrors the ratio of the causative and anticausative morphological marking across a representative sample of 21 languages (Haspelmath 1993; Samardžić and Merlo 2010; Samardžić and Merlo 2012; Samardžić 2013) and Samardžić and Merlo (to appear)[11].

This result has recently been replicated for seven languages (English, Japanese, Maltese, Romanian, Russian, Swahili, and Turkish) in a different study and for Spanish and French (Haspelmath et al. 2014; Heidinger 2015). Haspelmath et al. (2014) give a different interpretation of the results. Samardžić and Merlo argue that "likelihood of external causation" is the common hidden cause of both the morphological marking distributions across the typology and the frequency within a language, and that it is a formal lexical property anchored in the semantics of the verb. The lexical property is probabilistic and the observed frequencies of use are the surface expression of this probability. Haspelmath and colleagues instead put forth a functional model in which the variables are all fully observed, and there is no abstract underlying common cause. When the causative is marked morphologically, it is the less frequent, and the anticausative the more frequent. Hence their ratio of morphological marking is going to be correlated with the ratio of causative/anticausative use. In their view, shorter, unmarked forms are more frequent, while marked, longer forms are less frequent because this pairing of frequency and form is optimal and gives rise to the most efficient code. The correlation between corpus frequency and morphological marking is therefore simply and fully explained by communicative pressure. In this functional explanation, use and communicative needs, in this case efficiency of communication, shape the forms of the linguistic units, in this case the morphological encodings of causatives. This explanation however leaves some important aspects of the results unexplained. Haspelmath and colleagues find a relation between form and function only in the aggregated data, and not in the statistics of the individual languages. If form and function were really directly correlated to satisfy communicative pressure, we should find this correlation for each individual language, the actual means of communication.

Further confirmation of the alternative explanation based on the likelihood of a hidden variable is instead confirmed by an interesting result obtained using a parallel corpus (Samardžić 2013, Samardžić and Merlo, to appear). Most typological studies use separate grammar or corpus-resources for each individual language, thereby being limited to cross-linguistic observations at the level of the lexical entries of the verb

11 The data concerning the 21 languages are taken from Haspelmath (1993). The languages are Russian, Lithuanian, German, English, French, Rumanian, Greek, Armenian, Hindi-Urdu, Finnish, Hungarian, Udmurt, Arabic, Hebrew, Turkish, Khalkha Mongolian, Lezgian, Georgian, Swahili, Indonesian, Japanese. Haspelmath explains that the choice of the language sample was dictated by practical considerations.

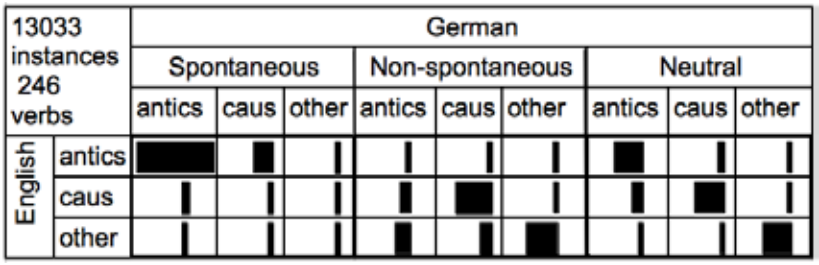


Figure 3
Effect of spontaneity variable in a parallel corpus: spontaneous = low likelihood; unspontaneous = high likelihood; neutral = mid likelihood.

(the type level). A parallel corpus, instead, supports investigations of cross-linguistic correspondence of individual verb occurrences, the token level[12]. This allows one to study how the quantitative lexical properties and the quantitative structural properties interact in a controlled, but naturalistic setting. Parallel corpora control for pragmatic and semantic factors as the two languages are translations of each other, but are naturalistic, and have a tendency to maintain parallel structures in translation. Any change in structural parallelism is then the indication of deeper differences.

It is found that the verb's likelihood of external causation modulates the choice of syntactic construction in the target language: for those sentence pairs where the verb is congruent with the construction (low likelihood of external causation in an anticausative construction, for example), the construction in the target language is the same as the construction in the source language. When the lexical property and the constructions are not congruent —low likelihood of external causation in a causative construction, for example—, then there is a tendency to regularise and shift, in the target, to the congruent construction. Figure 3 shows the joint distributions of the translations in three syntactic constructions, disaggreagted by high, low, and mid values of likelihood of external causation. This finding confirms the existence of an underlying component of meaning and supports models where the cross-linguistic distribution of morphological markings and the choice within a language of the causative alternant are determined by this same component. Frequency is not related here to structural or derivational complexity, but directly to an underlying probabilistic component in the lexical entry of the verbal root. This probabilistic component surfaces as frequency differentials. Across languages, this probabilistic component can match or mismatch the structure and generates regularisation effects.

We have here a very different kind of theory of frequency distributions from the complexity explanations proposed in the two previous sections. While the first case study shows that markedness, defined in terms of complexity, and frequency cannot be

12 Notice that we are using here the notion of type frequency in its simplest corpus-linguistic sense, as the marginal sum of the counts tallied over all the tokens of a given type in a classification. In usage-based explanations a slightly different concept has been assigned the same term, type frequency, to describe the range of possible types that can occur in a given position in a sentence, or slot in a schema. For example, the cardinality of the set of verbs that can occur with the ending $-ed$ in the apst tense. In usage-based accounts, type frequency is a measure of productivity, and it refers to a distribution over types. See for example (Bybee and Beckner 2009).

fully equated, this case study argues, instead, that frequency can be generated from a hidden probabilistic component that fully explains it.

3. Discussion and Conclusions

In this work, we have taken frequencies to be the linguistic object of study and have presented two different kinds of theories of frequency differentials. We have discussed two kinds of independent explanatory variables: underlying complexity and underlying probabilities. The complexity-based theories come in different flavours, based on different views of complexity: complexity can be the result of derivation steps or costs of operations, or both, as in Cinque's explanation of Universal 20. Complexity can also be seen as the violations of some high-level principle of grammar construction, like harmony or symmetry in Dryer's explanation of typology differentials. Dependency Length Minimisation is a locality or distance-based theory of complexity, where a precise and predictive calculation of complexity is anchored in the exact structures of the language. The proposal for causatives augments well-established lexical semantic approaches to verb meaning based on underlying root meaning decomposition (Hale and Keyser 1993; Alexiadou 2010). It takes the fundamental insight of these approaches —that components of meaning surface as different structures— and shows that lexical meaning components can be probabilistic in nature and give rise to the different observed frequencies of syntactic alternations.

Beyond the survey of the three case studies and their technical contributions, this paper also means to offer a novel view on corpus-driven theoretical linguistics. While the use of corpora is common-place in computational linguistics and the use of quantitative and statistical approaches is inescapable in psycholinguistics or experimental methods, the use of corpora and computational methods to address syntactic questions of a formal nature is not wide-spread. It is also not trivial. This paper offers also proposals of new methodologies. In so doing, it demonstrates that computational models provide a sound formal methodology that yields theoretical insights. It shows that large-scale corpora, both monolingual and parallel, are useful sources of cross-linguistic information on core syntactic questions and fine-grained investigations. Far from being theoretically irrelevant and uninformative, quantitative computational approaches exploit complex data resources and inspire articulated quantitative explanations that ultimately lead to better empirical coverage and simpler theories.

Acknowledgments

I gratefully thank the Swiss NSF which has funded, under grants 122643 and 144362, the great collaborators who have worked with me on this research program: Kristina Gulordava, Sarah Ouwayda, Tanja Samardžić. The members of the Computational Learning and Computational Linguistics group have given comments at different stages of this work. Thanks to the participants and organisers of Clic-it 2015 in Trento and the TABU Dag in Groningen and two anonymous reviewers. I particularly thank Simonetta Montemagni and Roberto Basili for suggesting I write up this piece. Thanks also to Guglielmo Cinque for providing the data on Universal 20. All remaining errors are my own.

References

Abels, Klaus and Ad Neeleman. 2009. Universal 20 without the LCA. In A. Gavarro J.M. Brucart and J. Sola, editors, *Merging Features: Computation, Interpretation, and Acquisition*. Oxford

University Press, Oxford/New York, pages 60–79.
Abend, Omri, Roi Reichart, and Ari Rappoport. 2008. A supervised algorithm for verb disambiguation into verbnet classes. In *Proceedings of the 22Nd International Conference on Computational Linguistics - Volume 1*, COLING '08, pages 9–16.
Alexiadou, Artemis. 2010. On the morpho-syntax of (anti-) causative verbs. In M. Rappaport-Hovav et al., editor, *Syntax, lexical semantics and event structure*. OUP, pages 177–203.
Baker, Mark C. 2002. *The atoms of language: The mind's hidden rules of grammar*. Basic books.
Baroni, Marco and Alessandro Lenci. 2010. Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics*, 36(4):673–721.
Bresnan, Joan. 2007. Is syntactic knowledge probabilistic? experiments with the english dative alternation. *Roots: Linguistics in search of its evidential base*, pages 75–96.
Bresnan, Joan, Anna Cueni, Tatiana Nikitina, and Harald Baayen. 2007. Predicting the dative alternation. In G. Boume, I. Kraemer, and J. Zwarts, editors, *Cognitive Foundations of Interpretation*. Royal Netherlands Academy of Science, Amsterdam, pages 69–94.
Bresnan, Joan, Shipra Dingare, and Christopher D Manning. 2001. Soft constraints mirror hard constraints: Voice and person in english and lummi. In *Proceedings of the LFG01 Conference*, pages 13–32.
Buchholz, Sabine and Erwin Marsi. 2006. CoNLL-X Shared Task on Multilingual Dependency Parsing. In *Proceedings of the Tenth Conference on Computational Natural Language Learning*, CoNLL-X '06, pages 149–164, Stroudsburg, PA, USA. Association for Computational Linguistics.
Bybee, Joan. 2007. *Frequency of Use and the Organisation of Language*. Oxford University Press.
Bybee, Joan L. and Clay Beckner. 2009. Usage-based theory. In Edited by Bernd Heine and Heiko Narrog, editors, *The Oxford Handbook of Linguistic Analysis*. Oxford University Press, pages 827–856. DOI: 10.1093/oxfordhb/9780199544004.013.0032.
Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA.
Cinque, Guglielmo. 2005. Deriving Greenberg's universal 20 and its exceptions. *Linguistic Inquiry*, 36(3):315–332.
Cinque, Guglielmo. 2013. *Typological Studies: Word Order and Relative Clauses*. Routledge, New York/London.
Clegg, Andrew B and Adrian J Shepherd. 2007. Benchmarking natural-language parsers for biological applications using dependency graphs. *BMC bioinformatics*, 8(1):24.
Cysouw, Michael. 2010. Dealing with diversity: towards an explanation of NP word order frequencies. *Linguistic Typology*, 14(2):253–287.
Demberg, Vera and Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2):193–210.
Dryer, Matthew. 2006. The order demonstrative, numeral, adjective and noun: an alternative to Cinque. http://exadmin.matita.net/uploads/pagine/1898313034_cinqueH09.pdf.
Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language*, 68:81–138.
Dryer, Matthew S. 2009. The branching direction theory of word order correlations revisited. In *Universals of language today*. Springer, pages 185–207.
Dryer, Matthew S., 2013. *Determining Dominant Word Order*. Max Planck Institute for Evolutionary Anthropology, Leipzig.
Fedzechkina, Maryia, T Florian Jaeger, and Elissa L Newport. 2012. Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences*, 109(44):17897–17902.
Futrell, Richard, Kyle Mahowald, and Edward Gibson. 2015. Large-Scale Evidence of Dependency Length Minimization in 37 Languages. *Proceedings of the National Academy of Sciences of the United States of America*.
Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1):1–76.
Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. *Image, language, brain*, pages 95–126.
Gildea, Daniel and David Temperley. 2010. Do Grammars Minimize Dependency Length? *Cognitive Science*, 34(2):286–310.
Greenberg, Joseph H. 1966. *Language Universals*. Mouton, The Hague, Paris.
Gulordava, Kristina and Paola Merlo. 2015a. Diachronic trends in word order freedom and dependency length in dependency-annotated corpora of latin and ancient greek. In *Proceedings*

of the International Conference on Dependency Linguistics (DepLing'15), Uppsala, Sweden, August.
Gulordava, Kristina and Paola Merlo. 2015b. Structural and lexical factors in adjective placement in complex noun phrases across romance languages. In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning*, pages 247–257, Beijing, China, July.
Gulordava, Kristina, Paola Merlo, and Benoit Crabbé. 2015. Dependency length minimisation effects in short spans: a large-scale analysis of adjective placement in complex noun phrases. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 477–482, Beijing, China, July.
Hajič, Jan, Massimiliano Ciaramita, Richard Johansson, Daisuke Kawahara, Maria Antònia Martí, Lluís Màrquez, Adam Meyers, Joakim Nivre, Sebastian Padó, Jan Štěpánek, Pavel Straňák, Mihai Surdeanu, Nianwen Xue, and Yi Zhang. 2009. The CoNLL-2009 shared task: syntactic and semantic dependencies in multiple languages. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task*, CoNLL '09, pages 1–18, Stroudsburg, PA, USA. Association for Computational Linguistics.
Hale, Ken and Jay Keyser. 1993. On argument structure and the lexical representation of syntactic relations. In K. Hale and J. Keyser, editors, *The View from Building 20*. MIT Press, pages 53–110.
Haspelmath, Martin. 1993. More on typology of inchoative/causative verb alternations. In Bernard Comrie and Maria Polinsky, editors, *Causatives and transitivity*, volume 23. John Benjamins Publishing Co., Amsterdam/Philadelphia, pages 87–121.
Haspelmath, Martin. 2006. Against markedness (and what to replace it with). *Journal of linguistics*, 42(1):25–70.
Haspelmath, Martin, Andreea Calude, Michael Spagnol, Heiko Narrog, and Elif Bamyac. 2014. Coding causal-noncausal verb alternations: a form-frequency correspondence explanation. *Journal of Linguistics*, 50(3):587–625.
Hawkins, John A. 1994. *A Performance Theory of Order and Constituency*. Cambridge University Press, Cambridge.
Hawkins, John A. 2004. *Efficiency and Complexity in Grammars*. Oxford University Press, Oxford.
Heidinger, Steffen. 2015. Causalness and the encoding of the causative–anticausative alternation in french and spanish. *Journal of Linguistics.*, 51(3):562–594.
Levin, Beth. 1993. *English Verb Classes and Alternations*. University of Chicago Press, Chicago, IL.
Lohse, Barbara, John A. Hawkins, and Thomas Wasow. 2004. Domain minimization in english verb-particle constructions. *Language*, pages 238–261.
McDonald, Ryan, Joakim Nivre, Yvonne Quirmbach-Brundage, Yoav Goldberg, Dipanjan Das, Kuzman Ganchev, Keith Hall, Slav Petrov, Hao Zhang, Oscar Täckström, Claudia Bedini, Núria Bertomeu Castelló, and Jungmee Lee. 2013. Universal dependency annotation for multilingual parsing. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 92–97, Sofia, Bulgaria, August. Association for Computational Linguistics.
Merlo, Paola. 1994. A corpus-based analysis of verb continuation frequencies for syntactic processing. *Journal of Psycholinguistic Research*, 23(6):435–457.
Merlo, Paola. 2015. Predicting word order universals. *Journal of Language Modelling*, 3(2):317–344.
Merlo, Paola and Sarah Ouwayda. Movement and structure effects on universal 20 word order frequencies: a quantitative study. submitted.
Merlo, Paola and Suzanne Stevenson. Structure and frequency in verb classification. In *Volume collecting papers from the 30th Incontro di Grammatica Generativa*.
Merlo, Paola and Suzanne Stevenson. 2001. Automatic verb classification based on statistical distributions of argument structure. *Computational Linguistics*, 27(3):373–408.
Merlo, Paola, Suzanne Stevenson, Vivian Tsang, and Gianluca Allaria. 2002. A multi-lingual paradigm for automatic verb classification. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02)*, pages 207–214, Philadelphia, PA.
Miller, George and Kathryn McKean. 1964. A chronometric study of some relations between sentences. *Quarterly Journal of Experimental Psychology*, 16:297–308.
Moravcsik, Edith and Jessica Wirth. 1983. Markedness – an Overview. In Fred Eckman, Edith Moravcsik, and Jessica Wirth, editors, *Markedness*. Plenum Press, New York,NY, pages 1–13.
Palmer, Martha, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31:71–105.
Reinhart, Tanja. 2002. The theta system — An overview. *Theoretical linguistics*, 28:229–290.
Rizzi, Luigi. 1990. *Relativized Minimality*. MIT Press, Cambridge, MA.

Samardžić, Tanja. 2013. *Dynamics, causation, duration in the predicate-argument structure of verbs: A computational approach based on parallel corpora*. Ph.D. thesis, University of Geneva.
Samardžić, Tanja and Paola Merlo. 2010. Cross-lingual variation of light verb constructions: Using parallel corpora and automatic alignment for linguistic research. In *Proceedings of the 2010 Workshop on NLP and Linguistics: Finding the Common Ground*, pages 52–60, Uppsala, Sweden, July. Association for Computational Linguistics.
Samardžić, Tanja and Paola Merlo. forthcoming. Probability of external causation: an empirical account of cross-linguistic variation in lexical causatives. *submitted*.
Samardžić, Tanja and Paola Merlo. 2012. The meaning of lexical causatives in cross-linguistic variation. *Linguistic Issues in Language Technology*, 7(12):1–14.
Schaefer, Florian. 2009. The causative alternation. *Language and linguistics compass*, 3:641–681.
Schulte im Walde, Sabine. 2006. Experiments on the automatic induction of german semantic verb classes. *Computational Linguistics*, 32(2):159–194.
Scott, R. M. and Cynthia Fisher. 2009. Two-year-olds use distributional cues to interpret transitivity-alternating verbs. *Language and Cognitive Processes*, 24(6):777–803.
Shlonsky, Ur. 2012. On some properties of nominals in Hebrew and Arabic, the construct state and the mechanisms of AGREE and MOVE. *Italian Journal of Linguistics*, 24(2):267–286.
Steddy, Sam and Vieri Samek-Lodovici. 2011. On the ungrammaticality of remnant movement in the derivation of greenberg's universal 20. *Linguistic Inquiry*, 42(3):445–469.
Steedman, Mark. 2011. Greenberg's 20th: The view from the long tail. unpublished manuscript, University of Edinburgh.
Temperley, David. 2007. Minimization of dependency length in written English. *Cognition*, 105(2):300–333.
Tily, Harry, Michael Frank, and Florian Jaeger. 2011. The learnability of constructed languages reflects typological patterns. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, pages 1364–1369.
Travis, Lisa. 1984. *Parameters and Effects of Word Order Variation*. Ph.D. thesis, MIT, Cambridge, MA.
Wasow, Thomas. 2002. *Postverbal Behavior*. CSLI Publications.
Yang, Charles. 2003. *Knowledge and Learning in Natural Language*. Oxford University Press.
Zipf, George Kingsley. 1949. Human behavior and the principle of least effort.