

# Dependency length minimisation effects in short spans: a large-scale analysis of adjective placement in complex noun phrases

**Kristina Gulordava**

University of Geneva

Kristina.Gulordava,

**Paola Merlo**

University of Geneva

Paola.Merlo@unige.ch

**Benoit Crabbé**

U. Paris Diderot/Inria/IUF

bcrabbe@  
univ-paris-diderot.fr

## Abstract

It has been extensively observed that languages minimise the distance between two related words. Dependency length minimisation effects are explained as a means to reduce memory load and for effective communication. In this paper, we ask whether they hold in typically short spans, such as noun phrases, which could be thought of being less subject to efficiency pressure. We demonstrate that minimisation does occur in short spans, but also that it is a complex effect: it is not only the length of the dependency that is at stake, but also the effect of the surrounding dependencies.

## 1 Introduction

One of the main goals in the study of language is to find explanations for those fundamental properties that are found in every human language. The observation that human languages appear to minimise the distance between any two related words – called the property of dependency length minimisation (DLM) — is a universal property that has been documented in sentence processing (Gibson, 1998; Hawkins, 1994; Hawkins, 2004; Demberg and Keller, 2008), in corpus properties of treebanks (Temperley, 2007; Futrell et al., 2015), in diachronic language change (Tily, 2010). Functional explanations have been proposed for this pervasive linguistic property. If speakers want to reduce memory load and maximise efficiency of processing, they will choose to produce and preferentially analyse constructions where words are linearised in such a way that minimises the total distance of related words.

The DLM principle can be stated as follows: if there exist possible alternative orderings of a phrase, the one with the shortest overall dependency length ( $DL$ ) is preferred. We measure the

length of a dependency as the number of words between the head and its dependent.

As an illustration, DLM principle is widely reported in the literature to explain the alternation of postverbal complements (Bresnan et al., 2007; Wasow, 2002). Consider, for example, the case 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  $DL_2$ .  $DL_1$  is  $DL(V-NP) + DL(V-PP) = |NP| + 1$ ;  $DL_2$  is  $DL(V-NP) + DL(V-PP) = |PP| + 1$ .<sup>1</sup> If  $DL_1$  is bigger than  $DL_2$ , then  $VP_2$  is preferred over  $VP_1$ , despite the non-canonical V-PP-NP order.

While DLM has been demonstrated on a large scale and explanations have been proposed based on human sentence processing facts in the verbal domain, it is not clear what the effects of DLM are in the more limited nominal domain. If the explanations are really rooted in memory and efficiency, will they still hold in phrases that might span only a few words?

In this paper, we look at the structural factors that play a role in adjective-noun word order alternations in Romance languages. We choose Romance languages because they show a good amount of variation, making studies of DLM meaningful. This would not be the case in English, for instance, as English has no variation of word order placement in the noun phrase. Adjective placement in Romance is often studied in connection with semantic and lexical properties of adjectives (Bouchard, 1998; Cinque, 2010). There exists, however, a body of work which shows that structural syntactic properties like the size of adjective phrase also affect the adjective position (Abeillé and Godard, 2000; Thuilier, 2012).

We demonstrate that, unlike results for the ver-

<sup>1</sup>The minimal dependency length is equal to one when the head and its dependent are adjacent.

	RightNP=Yes	RightNP=No
X=Left	$ \beta  -  \alpha $	$2 \beta  + 1$
X=Right	$-3 \alpha  - 2$	$-2 \alpha  - 1$

Table 1: Dependency length difference for different types of noun phrases. By convention, we always calculate  $DL_1 - DL_2$ .

bal domain, it is not only the length of the dependency that is at stake, but also the effect of the surrounding dependencies.

## 2 Dependency length minimisation in the noun phrase

In applying the general principle of DLM to the dependency structure of noun phrases, our goal is to test to what extent the DLM principle predicts the observed adjective-noun word order alternation pattern in relatively short spans.

Consider a prototypical noun phrase with an adjective phrase as a modifier. We assume two possible placements for an adjective phrase: postnominal and prenominal. To simplify, we concentrate on noun phrases with only one adjective modifier adjacent to the noun. The adjective modifier can be a complex phrase with both left and right dependents ( $\alpha$  and  $\beta$ , respectively, in Figure 1). The noun phrase can have parents and right modifiers (X and Y, respectively, in Figure 1). These alternative orderings yield different dependency lengths, as can be seen from Figure 1. By convention, we will always indicate the prenominal order as  $DL_1$ , and the postnominal order as  $DL_2$ . Their difference is always calculated as  $DL_1 - DL_2$ .

We consider all dependencies in a noun phrase and not only the length of the noun-adjective dependency. This is because we assume, as previously done, that DLM is global, and not a local, effect. Our analysis is a faithful interpretation of the very general DLM principle of Gildea and Temperley (2010) which is based on the overall dependency length of a sentence. We do not take other dependencies in the sentence into account, because their lengths are the same across  $DL_1$  and  $DL_2$ . The difference  $DL_1 - DL_2$  is therefore the difference between the overall dependency length of two sentences that differ only in their placement of one adjective.

The first panel, panel a, shows the case where the parent of the NP is on the left of it. The dependency length for the prenominal adjective struc-

ture is equal to  $DL_1 = d'_1 + d'_2 = (|\alpha| + |\beta| + 1) + |\beta|$  and for the postnominal adjective structure is  $DL_2 = d''_1 + d''_2 = |\alpha|$ . The difference between these lengths is  $2|\beta| + 1$ , which means that  $DL_1 > DL_2$  and suggests that the postnominal placement is always preferred.

Similarly, the second panel, panel b, in the figure shows how we calculate the dependency lengths when the parent of the NP is on its right. The difference of lengths is equal to  $-2|\alpha| - 1$ , yielding a preference for prenominal adjectives.

We also consider more complex noun phrases with at least one right dependent, which are very common in Romance languages (around 50% of noun phrases in our sample include, for instance, a complement, such as a relative clause). The third and fourth panels in Figure 1 illustrate the case where three dependencies should be taken into account. The calculations of these dependency lengths for the prenominal and postnominal alternatives yield the corresponding differences of  $|\beta| - |\alpha|$  (in the case of a left external dependency) and  $-3|\alpha| - 2$  (in the case of a right external dependency). These values are different from the dependency length differences for noun phrases without a right dependent (panel a and b). The comparison of the values, where RightNP=Yes is smaller than RightNP=No in both cases, suggests that the presence of a right dependent favours the prenominal placement of adjectives in comparison to the case of a simple noun phrase.

The differences in dependency lengths are summarized in Table 1. The expectations based on dependency length minimisation are as indicated in (1) below.

- (1) a. the presence of a left dependent of an adjective favours the adjective's prenominal placement;
- b. the presence of a right dependent of an adjective favours the adjective's postnominal placement;
- c. when the external dependency is leftwards, X = right, (for canonical subjects, for example), then the adjective is prenominal, because the difference is negative and it is a function of  $\alpha$ ;
- d. when the noun has a right dependent, the prenominal adjective position is more preferred than when there is no right dependent,

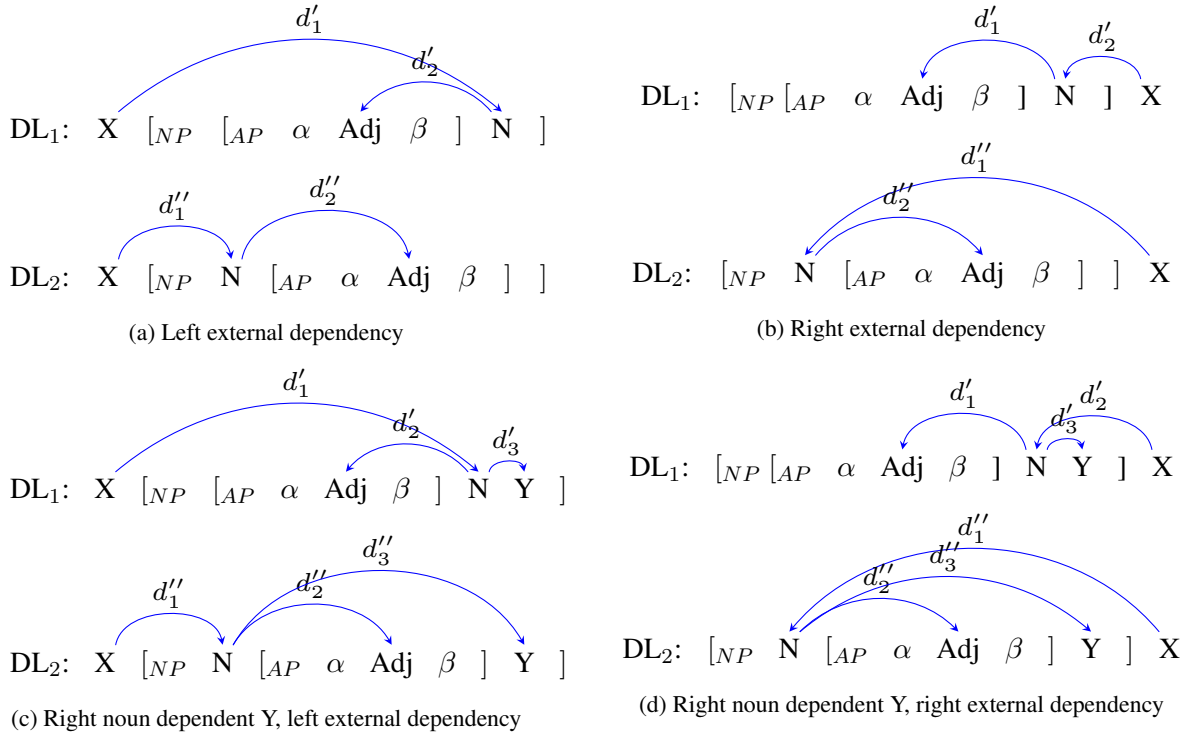


Figure 1: Noun phrase structure variants and the dependencies relevant for the DLM calculation.

as evinced by the fact that the RightNP = Yes column is always greater than the RightNP = No column.

The predictions (1a) and (1b) are formulated for an average case of adjective placement, across nouns phrases with different values of X and RightNP factors. Table 1 shows that for each combination of these context factors the weight of  $\alpha$  is negative or zero and the weight of  $\beta$  is positive or zero. On average, therefore, we expect to see a negative effect of  $\alpha$  (1a) and a positive effect of  $\beta$  (1b).

We develop a model to test which of the fine-grained predictions derived from DLM are confirmed by the data provided by the dependency annotated corpora of five of the main Romance languages.

### 3 Identifying dependency minimisation factors

#### 3.1 Materials: Dependency treebanks

We use the dependency annotated corpora of five Romance languages: Catalan, Spanish, Italian (Hajič et al., 2009), French (Agić et al., 2015), and Portuguese (Buchholz and Marsi, 2006).

We use part-of-speech information and dependency arcs from the gold annotation to extract

noun phrases containing adjectives. Specifically, we first convert all treebanks to coarse universal part-of-speech tags, using existing conventional mappings from the original tagset to the universal tagset (Petrov et al., 2012). We then identify all adjectives (tagged using the universal PoS tag ‘ADJ’) whose dependency head is a noun (tagged using the universal PoS tag ‘NOUN’). In addition, we recover all elements of the noun phrase rooted in this noun, that is, its dependency subtree. For all languages where this information is available, we extract lemmas of adjective and noun tokens which are the features in our analysis. The only treebank without lemma annotation is French, for which we extract token forms.<sup>2</sup> We extract a total of around 64’000 instances of adjectives in noun phrases, ranging from 2’800 for Italian to 20’000 for Spanish.

The data present a substantial amount of variation in the placement of the adjective: the ratio of postnominal adjectives ranges from around 65%

<sup>2</sup>During preprocessing, we exclude all adjectives and nouns with non-lower case and non-alphabetic symbols which can include common names, compounds (in Spanish and Catalan treebanks), and English borrowings. In addition, we leave out noun phrases which have their elements separated by punctuation (for example, commas or parentheses) to ensure that the placement of adjective is not affected by an unusual structure of the noun phrase.

for Italian to 78% for Catalan. Among all adjective types, at least 10% in each language are observed both prenominally and postnominally (ranging between 147 types for Italian and 445 types for Spanish).

### 3.2 Method: Mixed Effects models

We analyse the interactions of several dependency factors, using a logit mixed effect models (Bates et al., 2014). Mixed-effect logistic regression models (logit models) are a type of Generalized Linear Mixed Models with the logit link function and are designed for binomially distributed outcomes such as *Order* in our case.

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  $\beta$  and  $\gamma$  are the corresponding weights of the effects.

$$(2) \quad y = X\beta + Z\gamma + \epsilon$$

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

$$(3) \quad Order = \frac{1}{1 + \exp^{-y}}$$

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

### 3.3 Factors

We define and test the following factors, corresponding to the factors illustrated in Figure 1 and example (1), represented as binary or real-valued variables:

- *LeftAP* - the cumulative length (in words) of all left dependents of the adjective, indicated as  $\alpha$  in Figure 1;
- *RightAP* - the cumulative length (in words) of all right dependents of the adjective, indicated as  $\beta$  in Figure 1;
- *RightNP* - the indicator variable representing the presence ( $RightNP = 1$ ) or absence ( $RightNP = 0$ ) of the right dependent of the noun, indicated as  $Y$  in Figure 1;
- *ExtDep* - the direction of the arc from the noun to its parent  $X$ , an indicator variable.  $ExtDep = 0$  when  $X$  is on the left of the noun,  $ExtDep = 1$  when  $X$  is on the right.

Predictor	$\beta$	SE	Z	p
Intercept	-0.17	(0.117)	-1.42	0.16
LeftAP	2.21	(0.101)	21.91	< .001
RightAP	0.76	(0.054)	14.08	< .001
ExtDep	-0.06	(0.071)	-0.85	0.40
RightNP	-0.77	(0.050)	-15.34	< .001

Random effects	Var
Adjective	1.989
Language	0.024

Table 2: Summary of the fixed and random effects in the mixed-effects logit model ( $N = 15842$ ), shown in (4).

In addition, to account for lexical variation, we include adjective lemmas (for French, we include tokens) as grouping variables introducing random effects. For example, the instances of adjective-noun order for a particular adjective will share the same weight value  $\gamma$  for the adjective variable, but across different adjectives this value can vary.<sup>3</sup>

For a given example involving an adjective  $i$  and belonging to language  $j$ , the linear component of the model is shown in (4).

$$(4) \quad y_{ij} = LeftAP \cdot \beta_{LAP} + RightAP \cdot \beta_{RAP} + \\ + RightNP \cdot \beta_{RNP} + ExtDep \cdot \beta_{ED} + \\ + \gamma_{Adj_i} + \gamma_{Lang_j}$$

By fitting the logit mixed-effect model to our dataset, we find the fixed and random effects coefficients which best explain the data. To show that a factor has a statistically significant effect on adjective placement, we must show that its fixed effect coefficient is significantly different from zero.

### 3.4 Results

The logit mixed-effects model fitted to our data, shown in (4), reveals the following picture (Table 2).

Both the *LeftAP* and *RightAP* factors favour postnominal placement ( $\beta_{LAP} = 2.21$ ,  $\beta_{RAP} = 0.76$ ,  $p < 0.001$ ), however there are important differences between the two.

<sup>3</sup>We include only random intercepts because the size of the data is not sufficient to estimate the slope variables. In addition, for a robust estimation of the random effects, we include in our dataset only adjectives that are observed both prenominally and postnominally.

LeftAP shows a complex behavior. When LeftAP is equal to one, it favors (slightly) the prenominal placement and when LeftAP is greater than one, it favors the postnominal placement. This result suggests that the adjective can behave differently depending on the size or type of its left periphery. For the moment it is not clear if the difference is due to length or type, as LeftAP of length one are almost always adverbs. It is important to notice that the results for LeftAP then do not entirely pattern with the predictions of dependency length minimisation, shown in (1a).

The RightAP factor shows a consistent postnominal preference, positively correlated to its length. Consequently, we can say that the RightAP is a stronger indicator of the postnominal placement than LeftAP, in agreement with the previously observed ordering patterns of adjective phrases (Abeillé and Godard, 2000) and the DLM prediction.

The external dependency factor is not significant ( $p > 0.1$ ). Moreover, the log likelihood ratio between the full model and the model without *ExtDep* is  $\chi^2$  distributed with 1 degree of freedom with  $\chi^2 = 3.8, p = 0.052$ . This comparison confirms that the introduction of the external dependency does not help predicting the Order. At first sight, this result suggests that this dependency is not subject to the minimisation principle. A plausible explanation claims that only the dependencies between the head and the *edge* of the dependent phrase are minimised (Hawkins, 1994). In Romance languages, the majority of the noun phrases take an article which unambiguously defines the left edge of the noun phrase. There is no need therefore to minimize the external dependency to the noun, since the noun phrase can be entirely predicted based on its left corner.

The RightNP factor is significant in the fitted model ( $\beta_{RNP} = -0.77, p < 0.001$ ).<sup>4</sup> The presence of a noun dependent on the right of the noun favours a prenominal placement, as predicted by DLM (1d). This is a result which, to our knowledge, was not previously observed in the literature, and that clearly answers our initial question, confirming that DLM also applies to very short spans. A much more detailed study of the lexical and structural properties of this effect is developed in

<sup>4</sup>A log-likelihood test of the model including RightAP, LeftAP and RightNP factors compared to the model including only RightAP and LeftAP factors yields  $\chi^2 = 107$  and  $p < .001$ .

(Gulordava and Merlo, 2015).

## 4 Conclusion

In this paper, we have developed a model of dependency length minimisation in the noun phrase and shown subtle interactions among its subcomponents. We show that most of DLM predictions are confirmed, and that DLM also apply to short spans. The fact that DLM effects also hold in such short spans casts doubts, in our opinion, on the grounding of this effect in memory limitations. The subtle interactions also raise questions on the exact definition of what dependencies are minimised and to what extent a given dependency annotation captures these distinctions, questions that we reserve for future work.

## Acknowledgements

We gratefully acknowledge the partial funding of this work by the Swiss National Science Foundation, under grant 144362 to the second author. Thanks also to the Labex EFL (ANR-10-LABX-0083 EFL) in Paris for supporting the visit of the first author.

## References

- Anne Abeillé and Daniele Godard. 2000. French word order and lexical weight. In Robert D. Borsley, editor, *The nature and function of Syntactic Categories*, volume 32 of *Syntax and Semantics*, pages 325–360. BRILL.
- Željko Agić, Maria Jesus Aranzabe, Aitziber Atutxa, Cristina Bosco, Jinho Choi, Marie-Catherine de Marneffe, Timothy Dozat, Richárd Farkas, Jennifer Foster, Filip Ginter, Iakes Goenaga, Koldo Gojenola, Yoav Goldberg, Jan Hajič, Anders Trærup Johannsen, Jenna Kanerva, Juha Kuokkala, Veronika Laippala, Alessandro Lenci, Krister Lindén, Nikola Ljubešić, Teresa Lynn, Christopher Manning, Héctor Alonso Martínez, Ryan McDonald, Anna Missilä, Simonetta Montemagni, Joakim Nivre, Hanna Nurmi, Petya Osenova, Slav Petrov, Jussi Piitulainen, Barbara Plank, Prokopis Prokopidis, Sampo Pyysalo, Wolfgang Seeker, Mojgan Seraji, Natalia Silveira, Maria Simi, Kiril Simov, Aaron Smith, Reut Tsarfaty, Veronika Vincze, and Daniel Zeman. 2015. Universal dependencies 1.1.
- Douglas Bates, Martin Maechler, Ben Bolker, and Steven Walker, 2014. *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7.

- Denis Bouchard. 1998. The distribution and interpretation of adjectives in French: A consequence of Bare Phrase Structure. *Probus*, 10(2):139–184.
- Joan Bresnan, 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*, pages 69–94. Royal Netherlands Academy of Science, Amsterdam.
- Sabine Buchholz 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.
- Guglielmo Cinque. 2010. *The Syntax of Adjectives: A Comparative Study*. MIT Press.
- Vera Demberg and Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2):193–210, November.
- Richard Futrell, Kyle Mahowald, and Edward Gibson. 2015. Large-Scale Evidence of Dependency Length Minimization in 37 Languages. (Submitted to Proceedings of the National Academy of Sciences of the United States of America).
- Edward Gibson. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1):1–76.
- Daniel Gildea and David Temperley. 2010. Do Grammars Minimize Dependency Length? *Cognitive Science*, 34(2):286–310.
- Kristina Gulordava and Paola Merlo. 2015. 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 (CoNLL'15)*.
- Jan Hajič, Massimiliano Ciaramita, Richard Johanson, 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.
- John A Hawkins. 1994. *A performance theory of order and constituency*. Cambridge University Press, Cambridge.
- John A. Hawkins. 2004. *Efficiency and Complexity in Grammars*. Oxford linguistics. Oxford University Press, Oxford, UK.
- Slav Petrov, Dipanjan Das, and Ryan T. McDonald. 2012. A Universal Part-of-Speech Tagset. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 2089–2096, Istanbul, Turkey.
- David Temperley. 2007. Minimization of dependency length in written English. *Cognition*, 105(2):300–333.
- Juliette Thuilier. 2012. *Contraintes préférentielles et ordre des mots en français*. Ph.D. Thesis, Université Paris-Diderot - Paris VII, Sep.
- Harry Joel Tily. 2010. *The role of processing complexity in word order variation and change*. Ph.D. Thesis, Stanford University.
- Thomas Wasow. 2002. *Postverbal Behavior*. CSLI Publications.