

# 1 Argument Structure

The classification systems discussed in this section considered the argument structure of verbs. In general, every finite verb in English takes one or more arguments, with a subject argument being required in normal speech and writing. The Stanford NLP system marks the arguments of verbs using the dependencies shown in Table 1.1. In the majority of non-copular sentences, the governors of these dependencies are the core verbs. In copular sentences the governor is generally the subject complement (i.e. the argument generally appearing after the verb which is equated with the subject) though in the case of copular sentences with clausal subjects, the Stanford parser chooses the copula to be the governor. Figure 1.1 show examples of these dependencies.

**Table 1.1:** The dependencies used to identify verbal arguments.

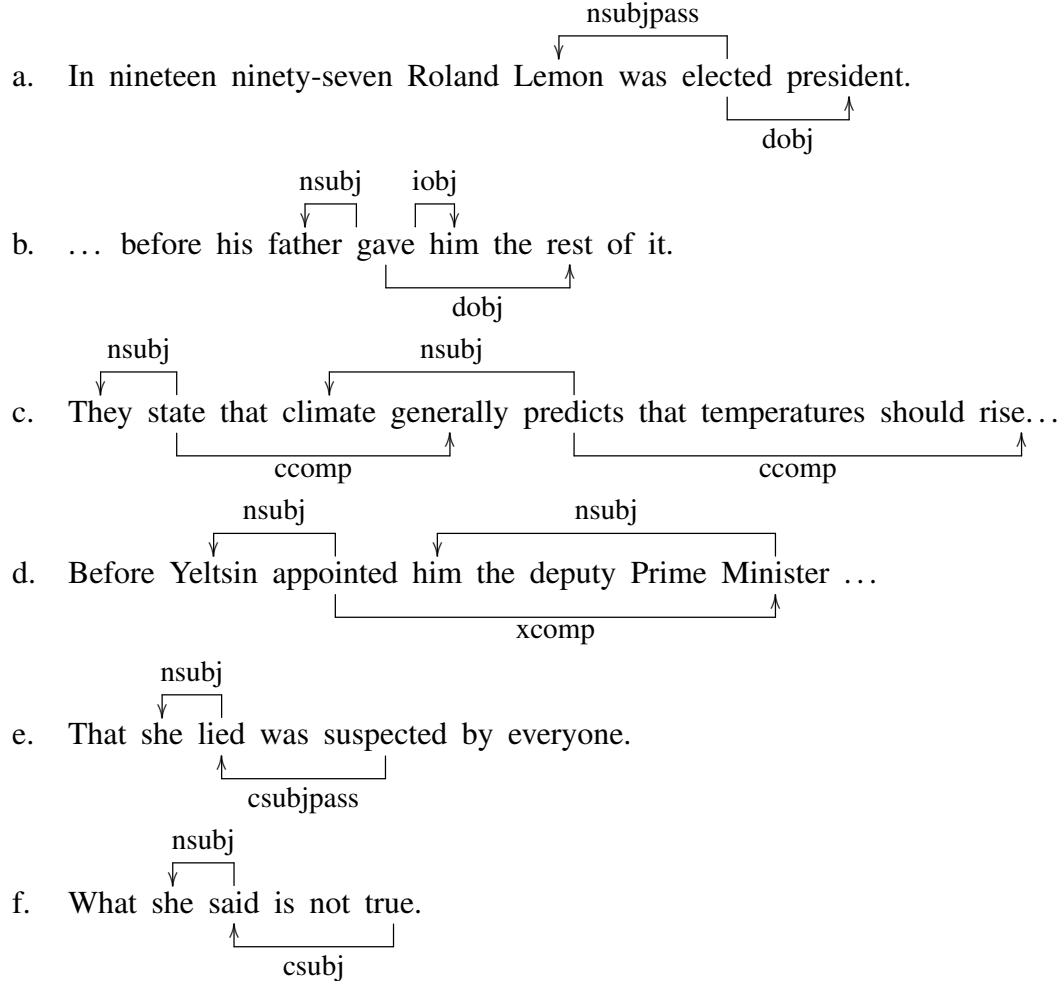
ccomp	Clausal Complement
csubj	Clausal Subject
csubjpass	Passive Clausal Subject
dobj	Direct Object
iobj	Indirect Object
nsubj	Nominal Subject
nsubjpass	Passive Nominal Subject
xcomp	Open Clausal Complement

Du Bois [2003] provides evidence that verbal argument structure may differ between native and nonnative speakers. That study explores when speakers choose to use lexical NPs rather than pronominal NPs as verbal arguments. Considering only native speakers but looking at a number of different languages, Du Bois finds that there is a very strong tendency for speakers to use no more than one lexical argument per finite clause (Du Bois' *One Lexical Argument Constraint*) and to avoid placing an argument in the subject role of transitive sentences (the *Non-Lexical A<sup>1</sup> Constraint*). He makes the case that these rules hold true for a number of world languages, at least in spoken form. However, in presenting data to show that several languages abide by the Non-Lexical A Constraint, he also shows

---

<sup>1</sup>Du Bois [2003] uses the letters **A**, **I**, and **O** to refer to the subject, indirect object, and direct object arguments of a transitive verb, and **S** to refer to the sole argument of an intransitive verb.

**Figure 1.1:** Examples of the Dependencies Listed in Table 1.1. *a*, native sample from ICE-CAN; *b,c,d*, native samples from MICUSP; *e,f* from de Marneffe and Manning [2008].



that there are large differences between languages in the likelihood of a lexical argument appearing in the direct object and intransitive subject roles [Du Bois 2003, Table 2.5]. His study shows that, based on his data, in English 21% of lexical arguments are found in intransitive clauses, versus 28% for Spanish, and 79% are found in direct object roles, versus 71% for Spanish. A  $\chi^2$  analysis shows that these differences are statistically significant considering his sample sizes ( $\chi^2 = 2.244$ ,  $df = 2$ ,  $\rho = 0.326$ ). As an aside, of the five languages for which data is presented in that table, English and Spanish are actually the most similar in lexical argument distribution. Of the three other languages considered, French, Hebrew, and Sakapultek, all show a greater probability of finding a lexical NP in the intran-

sitive subject role, and a lower probability of finding one in direct object role, than either Spanish or English. Seeing that there are differences between Spanish and English in how lexical arguments are distributed among the various argument roles, it is worth investigating whether L1-Spanish learners of English differ in their usage of these lexical arguments as compared to native speakers.

To investigate the feasibility of classifying language based on argument structure, I used Stanford dependency graphs to identify 18 different types of finite clauses: intransitives with and without clausal subjects, copular clauses with and without clausal subjects, simple transitives with and without clausal subjects and objects, ditransitives with and without clausal subjects, complex transitives with and without clausal subjects, passives of simple transitives with or without clausal subjects, and passives of complex transitives with or without clausal subjects and complements. Then for each of these 18 different types of clauses, all possible permutations of lexical and non-lexical arguments were considered, yielding 80 different attributes in total. This system, while providing a generous amount of data for the classifiers, proved to yield decision trees that were difficult to interpret linguistically. Fortunately, I found that from these attributes I could generate three much smaller and more coherent set of attributes which, when combined and used to train a classifier, provided comparable accuracy.

The first of these attribute sets consisted of just three arguments, corresponding to verb valency (i.e. the number of arguments). The values associated with these attributes were the percentage of all finite clauses which used that number of arguments. Figure 1.2 shows a C4.5 decision tree trained on these attributes. The next attribute set consisted of four attributes, named *zero*, *one*, *two*, *three* with corresponding values indicating what percentage of finite clauses had that number of lexical arguments. Figure 1.3 shows the C4.5 generated from these attributes. The last attribute set consisted of seven attributes corresponding to types of arguments. The values associated with these attributes were the percentage of lexical arguments were found in that type of argument. The seven types were intransitive,

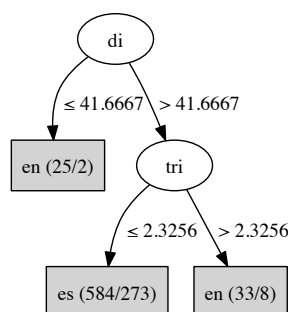
transitive, passive, and copular subjects, indirect and direct objects, and subject complements. This latter category included both the complement of copular clauses and what is usually the third argument in a complex transitive (e.g. *deputy Prime Minister* in Table 1.1).

To determine what was a lexical argument and what was not, I compiled a list of pronominal forms, largely taken from Celce-Murcia and Larsen-Freeman [1999, Ch. 16] and shown in Table 1.2, and considered any argument that did not match one of these forms to be lexical. By this standard, clausal arguments were always considered lexical arguments.

**Table 1.2:** Pronouns Used to Determine Non-Lexical Status of Arguments.

other	another	else	same	one	
this	that	these	those	what	
myself	yourself	herself	himself	itself	
ourselves	yourselves	themselves	oneself		
mine	yours	hers	his	ours	theirs
me	you	her	him	it	us
them	I	she	he	we	they

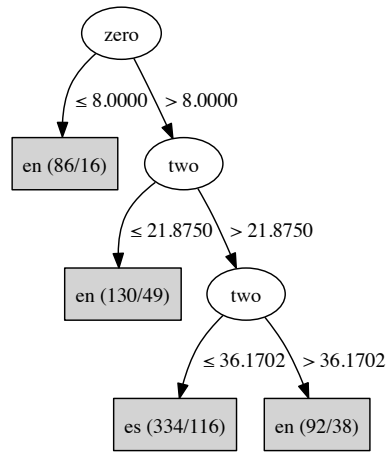
**Figure 1.2:** Verbal clause valency C4.5 decision tree. The three attributes have values equal to the percentage of finite verbal clauses which take one, two, or three arguments.



## References

CELCE-MURCIA, M. AND LARSEN-FREEMAN, D. 1999. *The Grammar Book, an ESL/EFL Teacher's Course* Second Ed. Heinle and Heinle Publishers.

**Figure 1.3:** Lexical argument quantity C4.5 decision tree. The four attributes have values equal to the percentage of finite verbal clauses with zero, one, two, or three lexical arguments.



DE MARNEFFE, M.-C. AND MANNING, C. D. 2008. Stanford typed dependencies manual.

DU BOIS, J. W. 2003. Discourse and grammar. In *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure*, M. Tomasello, Ed. Vol. 2. Lawrence Erlbaum Associates, Chapter 2, 47–87.

**Figure 1.4:** Lexical argument role C4.5 decision tree. Attributes have values indicating the percentage of all lexical arguments which are found in that role.

