

# Constituency and Dependency Parsers Evaluation

## *Evaluación de analizadores de constituyentes y de dependencias*

<b>Elisabet Comelles</b> Universitat de Barcelona Gran Via Corts Catalanes, 585 08007 Barcelona elicomelles@ub.edu	<b>Victoria Arranz</b> ELDA 55-57, rue Brillat Savarin 75013 Paris, FRANCE arranz@elda.org	<b>Irene Castellón</b> Universitat de Barcelona Gran Via Corts Catalanes, 585 08007 Barcelona icastellon@ub.edu
--	--	---

**Resumen:** El presente trabajo muestra la evaluación cuantitativa y cualitativa de un grupo de analizadores de constituyentes y de dependencias con el objetivo de ser usados en el desarrollo de una métrica automática basada en conocimiento para evaluar la salida de sistemas de traducción automática. Primero se describe la metodología seguida en ambos tipos de evaluación y a continuación se muestran los resultados obtenidos y las conclusiones alcanzadas.

**Palabras clave:** Analizador de dependencias, analizador sintáctico, evaluación cualitativa, evaluación cuantitativa, análisis de errores lingüísticos

**Abstract:** This work presents the quantitative and qualitative evaluation of a set of both constituency and dependency parsers which are to be used in the development of a knowledge-based automatic MT metric. Firstly, the methodology used in both types of evaluation is described; secondly, we show the results obtained, and finally we draw some conclusions.

**Keywords:** Constituency parser, dependency parser, qualitative evaluation, quantitative evaluation, linguistic errors analysis

## 1 Introduction

Nowadays, there is quite an important number of syntactic parsers available. When deciding which one to use it is important not only to take into account their processing speed, but also their quality in terms of linguistic analysis. We faced this problem when looking for both a constituency and a dependency parser that may be used in the development of a knowledge-based metric so as to automatically assess machine translation (MT) output. Current work on MT evaluation shows that purely statistical metrics such as BLEU do not do justice to MT performance (Hovy, 2007). A number of studies are currently trying to go beyond by looking into the linguistic aspects of translation, so as to judge its output in a more fair way (Liu and Gildea, 2005). The authors of this paper are working in this direction by studying the impact of syntactic and semantic information in the process (Giménez et al., 2010).

As the MT output had to correlate with human judgements, it was important to use high quality

tools which offered the best performance, and thus we decided to make a comparative evaluation among several parsers to see which one obtained the best results in terms of quality. Thus, the aim of this paper is to explain the methodology used in order to evaluate both constituency and dependency parsers, to show the results obtained and discuss the findings.

Several efforts have been made in order to deal with the evaluation of parsers. However, this is not a simple task and it has not been widely covered. Although there are several international competitions, campaigns and even gatherings focusing on the area (PASSAGE<sup>1</sup>, EVALITA<sup>2</sup>, workshop at Coling 2008<sup>3</sup> or CoNLL 2007<sup>4</sup> and 2009<sup>5</sup>) most of the shared tasks aimed at evaluation do not have continuity in time or they focus on single languages. Moreover, there are some restrictions in relation

<sup>1</sup> <http://atoll.inria.fr/passage/eval2.en.html>

<sup>2</sup> <http://evalita.fbk.eu/parsing.html>

<sup>3</sup> <http://lingo.stanford.edu/events/08/pe/>

<sup>4</sup> <http://www.cs.jhu.edu/EMNLP-CoNLL-2007/>

<sup>5</sup> <http://ufal.mff.cuni.cz/conll2009-st/>

to the type of parsers and the type of evaluation. For instance, in the CoNLL 2007 Shared Task (Nivre et al., 2007) where the performance of dependency parsers for the English language was assessed, most of the parsers tested were statistically-based, which meant that they were evaluated on the test part of the learning corpus. This fact allows the comparison between several parsers, but it does not grant their linguistic quality when dealing with tests they have not been trained on. Furthermore, results obtained in this type of competitions respond only to general statistical measures, such as ParsEval (Black et al., 1991), precision and recall, etc. In other words, the evaluation performed is a quantitative evaluation but it does not analyse the linguistic errors produced. Therefore, although more time-consuming and expensive, we considered that it was highly recommendable to evaluate parsers by means of human judgments, which allow to study the linguistic quality and to look into further improvement or adaptations.

This paper is structured as follows: Section 2 introduces the parsers evaluated. The Methodology section explains the two types of evaluation performed and the items we focused on depending on the type of parser assessed. In the Results and Discussion section we show the results obtained in our evaluation and we analyse the type of errors found. Finally, we draw some conclusions on the results obtained.

## 2 Parsers Evaluated

As mentioned in the introduction, nowadays there is a wide range of syntactic parsers which can be used in order to process natural language. When studying the possible candidates and due to the use that we wanted to give them, we decided to test both constituency and dependency parsers. The choice for these specific parsers was mostly based on availability, as well as ease to use due to their linguistic features.

Regarding constituency parsers, we tested the Charniak-Johnson's Max-Ent reranking parser (Charniak and Johnson, 2005), a parser which uses a regularized MaxEnt reranker to select the best parse from the 50-best parses returned by a generative parsing model. In addition, we also evaluated the statistical natural language parser developed by Collins (Collins, 1999).

With regard to dependency analysis, we tested 5 parsers. It must be noticed that some of these parsers could provide both constituency and dependency output. However, as the dependency analysis is directly related to the constituency one we decided to assess them only as dependency parsers. Firstly, we assessed the Stanford parser (Klein and Manning, 2003), a statistical parser which performs both constituency and dependency analyses (De Marneffe, MacCartney and Manning, 2006). Secondly, we tested the DeSR dependency parser (Attardi, 2006), a statistically-based shift-reduce parser. Thirdly, we tried another statistically-based parser, the RASP system (Briscoe, Carroll and Watson, 2006). In addition we evaluated the MINIPAR parser (Linn, 1998), a broad-coverage parser for the English language. Finally, we tested the MALT parser (Nivre, 2006), a system for data-driven dependency parsing.

## 3 Methodology

In order to evaluate the several tools mentioned above, a corpus of evaluation containing 46 sentences (see Appendix 1) was compiled. These sentences were obtained from the *Clause Pattern DB* (Comelles et al., 2010), which consists of a total of 700 sentences extracted from a fiction novel corpus and classified according to their syntactic pattern. In line with other researchers who have aimed their parser evaluation to a construction-focused evaluation (Rimell, Clark and Steedman, 2009), the selection of the 46 evaluation sentences took into account their clause pattern so as to have a wide variety of syntactic phenomena to test. Thus, the following syntactic patterns were included: SV, SVObl, SVA, SVCs, SVOd, SVOiOd, SVOdObl, SVOdA, SVOdCo, SVX-CompI, SVOdX-CompI, SVCl-CompI, SV[pass]Obl[by], SV[pass]Obl and SV[pass]OdObl[by] (see Appendix 2 for further information on correspondences between abbreviations and syntactic functions). In addition, the internal structure of phrases was also considered, trying to select those sentences which show a more complex phrase structure.

Once the corpus of evaluation was selected, a couple of evaluations were performed:

- a quantitative evaluation which took into account the number of sentences which were parsed correctly, and

- a qualitative evaluation which focused on the type of mistakes performed by the parsers.

These two types of evaluation were done when both assessing the constituency parsers and the dependency parsers. However, due to the different nature and output provided by the parsers, the steps followed when assessing the quality of the former or the latter differ. Thus, when testing constituency parsers we took into account the following parameters:

- Parts-of-Speech (PoS). We checked that the PoS were correctly assigned. As will be discussed in the Results section, a failure in the correct identification of PoS affects the rest of the analysis.
- Phrases. We assessed that both the scope of the phrases and the phrase label were correctly identified.

Regarding the evaluation of dependency parsers, on the one hand, we considered both dependency relations and their label (syntactic functions); and on the other hand, as we did in the constituency parsing evaluation, we also took into account the correct identification of PoS and phrases. Moreover, when we checked the dependencies we took into account the following levels of analysis:

- Phrase level. We checked that phrasal dependencies were assigned correctly.
- Sentence level. We checked that verbal constituents were correctly parsed.

It is worth mentioning that due to the ultimate goal of our evaluation (using the parser towards the development of a MT evaluation metric), we were very strict when dealing with errors performed by the parsers in both constituency and dependency evaluations. Thus, when an error of any kind was detected the sentence where it occurred was considered incorrect.

#### 4 Results and Discussion

This sections presents the results obtained for both constituency and dependency evaluations. We first present a quantitative evaluation with the results obtained for each parser and later we move to the qualitative analysis and we present

those positive and negative cases which are worth discussing.

#### 4.1 Constituency Parsers

We tested a couple of constituency parsers: Charniak and Collins parsers. Once these tools have been evaluated, results show that Charniak parser performs better than Collins system. As shown in Table 1, the Charniak parser analyses 89% of the sentences correctly whereas the performance of the Collins parser is slightly worse, obtaining 80% of correctly analysed sentences.

Parser	Correct Sentences	Failed Sentences
Charniak	89%	11%
Collins	80%	20%

Table 1

When looking in detail into the wrong sentences, we observe that both parsers made some common mistakes, such as the following:

- Wrong PoS assignment. This is a basic and crucial error, because once the parser assigns a PoS to a specific word, this PoS is kept during the whole process, hence it conditions and influences the analysis. Therefore, whenever a word is assigned a wrong PoS, this error is dragged throughout the whole analysis, thus causing problems in the identification of phrases. An example of this type of error can be seen in a couple of sentences: “Tomorrow Patrick will drive some of them to the airport” and “I remember mum walloping him with the broomstick”. In the first sentence, both parsers failed to identify “Tomorrow” as an adverb. Instead, they analyse “Tomorrow” either as a proper noun and the chunk “Tomorrow Patrick” as a NP. In the second sentence, a couple of words are misidentified. Firstly, “mum” is analysed as an adjective (Charniak parser). Secondly, “walloping” is identified as a noun instead of a verb (Collins and Charniak parsers). These two wrong PoS are dragged along the analysis and the whole chunk “mum walloping” is analysed as a NP instead

of splitting it into a couple of phrases: a NP “mum” and a non-finite VP “walloping”.

- Wrong identification of phrases. It seems that both tools sometimes fail in delimiting the scope of a phrase, and thus consider as part of a phrase some elements which should belong to a different one. For instance, in the sentence “He wiped the bottle dry with a dishcloth”, the Charniak parser analyses “dry with a dishcloth” as a whole AdjP instead of splitting it into an AdjP “dry” and a PP “with a dishcloth”, being both of them dependents on the verb “wipe”. A similar example can be found in the output of the sentence “You should never have donated defective sperm in the first place” misanalysed by the Collins parser. In this analysis, the parser considers “defective sperm in the first place” as a NP instead of dividing this chunk into two different phrases: a NP “defective sperm” and a PP “in the first place”. This seems to be a clear case which should be solved by means of a dependency analysis because the above PP is not a postmodifier of the noun “sperm”, but a non-argument of the verb “donate”.

Therefore, it seems that the main problems detected are caused by a bad identification of the scope of the phrases and a wrong assignment of the PoS. It must be noticed that most of the errors committed by the Collins parser are due to the scope of the phrase because it tends to consider PPs as post-modifiers of nouns instead of single constituents depending on the verb.

## 4.2 Dependency Parsers

Regarding dependency parsing, we evaluated 5 different systems: the Stanford parser, the DeSR parser, the RASP parser, the MINIPAR and the MALT parser. As shown in Table 2, the parser which shows the best performance is the Stanford parser, which gets 67% of correct sentences. The MALT parser reaches a total of 63% of correct sentences, followed by the MINIPAR and the DeSR parser which get 56% and 54% of correct sentences, respectively. The parser that gets the worst results is the RASP parser which has 45% of correct sentences.

Parser	Correct Sentences	Failed Sentences
Stanford	67%	33%
DeSR	54%	46%
RASP	45%	55%
MINIPAR	56%	44%
MALT	63%	37%

Table 2

It is worth mentioning that these errors correspond to both link and label errors. Thus if a link was correctly assigned but the label was wrong, it was counted as an error. If we compare these results to the average results obtained in international competitions, such as CoNLL, we notice that our results are much lower. This may be caused by several facts:

- The domain of the corpus (i.e. fiction novels) used in our evaluation is highly different to the corpus used in international competitions.
- Statistical parsers have a better performance when they are tested on a similar corpus to the one they have been trained on.
- As explained in the Methodology section, we performed a very strict evaluation and we penalised whole sentences even if they just show one error.

A close analysis of the errors shows that they can be classified into the following categories:

- Wrong assignment of PoS. As occurred in the previous section on constituency parser evaluation, if the parser assigns a wrong PoS to a word, this error is dragged throughout the whole analysis. The same sentences that were used to exemplify this error in the previous section can be applied here. In fact, out of all the parsers evaluated, MALT was the only one that parsed correctly the sentence “I remember mum walloping him with the broomstick”. This system identified “mum” as a noun and “walloping” as a non-finite verb. Moreover, only a couple of parsers (MALT and RASP) were capable of

identifying “Tomorrow” as an adverb in the sentence “Tomorrow Patrick will drive some of them to the airport”. It seems that assigning a PoS when there is ambiguity is a difficult matter to handle.

- Identification of more than one head. Not all parsers were capable of assigning a single head to every sentence, as illustrated by the analysis provided by the DeSR parser of the sentence “We met Derek’s daughter, the only person who can replace her father”. This system could not manage to identify the head of the sentence, which should be the verb “meet”. However, it provided two possible heads, the finite verb form “met” and the noun “person”.
- Wrong identification of head. Sometimes the parsers cannot identify the head of the sentence correctly. Although, usually the head of a clause is the verb, some parsers assign this function to another word category, such as the noun, as can be exemplified by the analysis performed by the MALT parser of the following sentence: “Worries about possible harmful effects have grown”, where the noun “worries” is considered as the head of the sentence, although it seems apparent that the head should be assigned to the finite verb form “have”.
- Wrong dependencies:
  - o PP-attachment. One of the most common errors found in the parsers assessed is PP-attachment. It seems to be a difficult process for most of the parsers to disambiguate and decide whether a PP depends on the verb, therefore being dependent on the verb, or on the preceding noun, performing the function of a post-modifier. For instance, in the sentence “Fred and George did not have the potions with them in the bathroom”, where the PP “with them” should depend on the verb “have”, most of the parsers (DeSR, MINIPAR and Stanford systems) analysed it as a post-modifier of the noun “potions”, and thus attached it to this noun.

However, the noun was not the only category to which PPs were wrongly attached. In the sentence “They had talked briefly on Saturday”, where the PP “on Saturday” depends on the non-finite verb form “talked”, a couple of the parsers, MINIPAR and RASP, attached the PP to the adverb “briefly”.

- o NP-attachment. This type of attachment occurs in fewer cases than PP-attachment, but it is worth highlighting because it causes problems when dealing with syntactic functions. In a sentence such as “My grandmother gave the visitors cakes and hot coffee”, the NPs “the visitors” and “cakes and hot coffee” depend both on the verb “gave”, thus working as Oi and Od, respectively. However, the output provided by all the systems tested, with the exception of the Stanford parser, attached “the visitors” to the noun “cakes, and thus causing a wrong assignment of the syntactic functions. The reason for this attachment seems to be that most of the parsers get “visitors” as a noun pre-modifying “cakes”.
- o AdvP-attachment. As for the NP-attachment, this type of mistake is not as common as the PP-attachment, but it is still important and worth discussing. In the sentence “John played with the boy kindly and patiently”, where the coordinated AdvP “kindly and patiently” should depend on “played”, some analysers (MALT, DeSR and RASP) wrongly attached this AdvP to the noun “boy”.
- o XCompl-attachment. Finally, most parsers attached the X-Complement to the preceding noun, as exemplified in the output of the analysis performed by the MALT, RASP, MINIPAR and Stanford parsers

of the sentence “I will feel the child’s foot to see if any bones are broken”, where the X-Complement “to see if any bones are broken” depends on the noun “foot” instead of on the verb “feel”.

- Syntactic functions. As mentioned in the methodology section, when assessing the several parsers, we took into account not only the dependency relations but also the labels or syntactic functions assigned to those dependency relations. During the evaluation process we noticed that sometimes although the dependency was correctly assigned, the syntactic function provided was wrong. That is the case of the evaluation that both the Stanford and DeSR parsers did of the sentence “My children were watching TV last night when I phoned them”. Although they were capable of determining that both “TV” and “last night” were NPs which depended on the non-finite verb form “watching”, when dealing with the syntactic functions, they wrongly assigned Oi and Object/Od labels respectively. The reason for this error may be the identification of “last night” as a NP instead of an AdvP. Although grammatically speaking, both categories are acceptable, the former can cause more problems when providing the syntactic function, because, as occurred in this analysis, the systems tend to analyse the sequence VP + NP + NP as Verb + Oi + Od.

Once studied the results and analysed the mistakes, it must be said that a correct identification of the Part-of-Speech and the scope of the phrase, as well as a good assignment of the phrasal category is crucial in order to obtain a good analysis. If a parser fails in one of these processes, the error is dragged throughout the analysis and the result is a badly parsed sentence.

On the other hand, the parsers evaluated are quite good at dealing with complex structures, such as relative structures. When parsing a sentence containing a relative construction, such as “I bought the children the aeroplane they had seen in the window”, all parsers analyse “they had seen in the window” as a relative clause depending on “aeroplane”.

Regarding other complex structures, containing X-Complements or Clause-Complements, the performance varies depending on the parser. The MALT and Stanford parsers, for instance, are quite good at dealing with such structures, whereas the rest of parsers show more problems when analysing them.

## 5 Conclusion

In this paper we have presented a comparative evaluation among several constituency and dependency parsers so as to use the best system towards the development of an automatic MT evaluation metric. Thus, we have performed a manual evaluation in order to identify the linguistic errors made by the parsers and see whether such type of information could be reliably applied to the evaluation of MT output, which will be part of our next experiments.

We have provided results on this evaluation and we have focused on the most common types of linguistic errors made by the parsers. After a close analysis of the errors, we conclude that one of the most common and crucial errors made by both constituency and dependency parsers are related to the PoS and Phrasal category assignment accuracy, as well as the identification of the scope of a phrase. It is also worth noticing that most parsers analyse correctly complex structures such as relative clauses, whereas most of them fail in analysing simple structures such as SVOiOd sentences which contain simple phrases. This could be explained by the different domain and type of syntactic structures in our corpus as compared to that used to train statistical parsers.

Since our aim is to continue our work on the development of the knowledge-based metric, we intend to explore further how the linguistic problems reported here could affect the metric itself. Given that our evaluation was very strict in terms of not accepting any type of error and not considering any value averaging, as some other evaluations do, we will also look into the implications behind the different types of linguistic levels (PoS, phrase level) as some of them may not be as harmful for the final objective of the annotation.

## Acknowledgments

This work has been funded by the Spanish Government project KNOW, TIN2009-14715-C0403. We would also like to thank the

Teaching Innovation Group GDI-GReLiC for letting us use their *Clause Pattern DB*.

## **Bibliography**

- Attardi, Giuseppe. 2006. Experiments with a Multilanguage Non-Projective Dependency Parser. In *Proceedings of the Tenth Conference on Natural Language Learning*. Pages 166-170, New York (USA).
- Black, Ezra, S. Abney, D. Flickenger, C. Gdaniec, R. Grishman, P. Harrison, D. Hindle, R. Ingria, F. Jelinek, J. Klavans, M. Liberman, M. Marcus, S. Roukos, B. Santorini and T. Strzalkowski. 1991. A procedure for quantitatively comparing the syntactic coverage of English grammars. In *Proceedings of DARPA Speech and Natural Language Workshop*, Pages 306–311, California (USA).
- Briscoe, Ted, J. Carroll and R. Watson. 2006. The Second Release of the RASP System. In *Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions*. Pages 77-80, Sydney (Australia).
- Charniak, Eugene and M. Johnson. 2005. Coarse-to-fine n-best parsing and MaxEnt discriminative reranking. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL)*. Pages 173-180, Michigan (USA).
- Collins, Michael. 1999. Head-Driven Statistical Models for Natural Language Parsing. PhD Dissertation, University of Pennsylvania (USA).
- Comelles, Elisabet, N. J. Laso, I. Verdaguer and E. Giménez. 2010. Clause Pattern DB: a corpus-based tool. Forthcoming in *II Congreso Internacional de Lingüística de Corpus*. A Coruña (Spain).
- De Marneffe, Marie-Catherine, B. MacCartney and C. D. Manning. 2006. Generation Typed Dependency Parsers from Phrase Structure Parses. In *International Conference on Language Resources and Evaluation*. Genoa (Italy).
- Giménez, Jesús, Ll. Márquez, E. Comelles, I. Castellón, V. Arranz. 2010. Document-level Automatic MT Evaluation based on Discourse Representations. Forthcoming in *ACL 2010. Fifth Workshop on Statistical Machine Translation and Metrics* MATR. Uppsala (Sweden).
- Hovy, Edward. 2007. Investigating why BLEU penalizes non-statistical systems. In *Proceedings of the eleventh MT Summit*. Copenhagen (Denmark).
- Klein, Dan and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting of Association for Computational Linguistics*. Pages 423–430, Sapporo (Japan).
- Lin, Dekang. 1998. Dependency-based Evaluation of MINIPAR. *Proceedings of the Workshop at LREC'98 on The Evaluation of Parsing systems*. Granada (Spain).
- Liu, Ding and Daniel Gildea. 2005. Syntactic Features for Evaluation of Machine Translation. In *ACL 2005 Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pp. 25–32, Ann Arbor, MI (USA).
- Nivre, Joakim. 2006. *Inductive Dependency Parsing*. Springer Verlag, Dordrecht. The Netherlands.
- Nivre, Joakim, J. Hall, S. Kübler, R. McDonald, J. Nilsson, S. Riedel and D. Yuret. 2007. The CoNLL 2007 Shared Task on Dependency Parsing. *Proceedings of the 2007 Conference on Empirical Methods in Natural Language and Computational Natural Language Learning*. Pages 915-932, Prague (Czech Republic).
- Rimell, Laura, S. Clark, and M. Steedman. 2009. Unbounded Dependency Recovery for Parser Evaluation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-09)*. Pages 813-821, Singapore.

## A Appendix 1: Evaluation Set

Set of evaluated sentences:

Evaluated Sentences
The country's foremost humorist died last night.
Worries about possible harmful effects have grown.
John played with the boy kindly and patiently.
They had talked briefly on Saturday.
The youngest daughter-in-law has cooked the dinner.
I'll feel the child's foot to see if any bones are broken.
Fred and George didn't have the potions with them in the bathroom.
A man named McBride has killed a convenience store clerk with a shotgun in Fresno in 1969.
The wizards behind Harry made noises of interests and excitement.
We met Derek's daughter, the only person who can replace her father.
Experts were smelling the funghi in order to identify them.
My children were watching TV last night when I phoned them.
Some of these weapons are donated to the state firearms lab.
The squatters were defended by a group of solicitors.
I was given this watch by my father.
That small room had probably once been a powder room.
She had been secretly proud of her calmed, controlled behaviour.
The trouble was that they didn't agree with the headmaster.
The birthmark over her eye was getting darker and darker.
The politician's speech convinced the audience.
You never should have donated defective sperm in the first place.
Tomorrow Patrick will drive some of them to the airport.
He was running his fingers distractedly over his backside.
The road of Destiny has taken us to those early accidents of fortune.
A number of things have contributed to that situation.
She was working on that account when she died.
He appointed you business guardian to his daughter and trustee to the vast fortune she inherited.
The afternoon has made the children quiet for a while.
He wiped the bottle dry with a dishcloth.
I bought the children the model aeroplane they had seen in the window.
My grandmother gave the visitors cakes and hot coffee.
I showed Harry the prophecy.

Jeremy gave me a look which advised me to shut my mouth.
The old professor donated all his books to the library.
White farmers promised freedom and a piece of bottom land to their slaves.
A commotion in the doorway behind them caused everyone to turn.
The school forbids the students to smoke in the classroom.
Her machine required you to say what kind of money you wanted.
I remember mum walloping him with the broomstick.
She knows when reports aren't being done on time.
David says that he would like for you to pay the expenses.
I think nobody inside this castle would have helped Black enter it.
Mark happened to be in Australia at that time.
The portrait is known to have been painted by an Italian.
She stopped writing to her English friends last year.
The police had promised to keep an eye on the place for the next few days.

## B Appendix 2: Correspondences

Correspondence between abbreviations and syntactic functions:

Abbreviation	Corresponding Syntactic Function
S	Subject
V	Verb
V[pass]	Passive verb
Oi	Indirect Object
Od	Direct Object
Obl	Oblique
Obl[by]	Oblique introduced by the preposition <i>by</i>
Cs	Subject Complement
Co	Object Complement
X-Compl	X-Complement / Catenative Construction
CI-Compl	Clause Complement