

Integrating NLP Tools in a Distributed Environment: A Case Study Chaining a Tagger with a Dependency Parser

Francesco Rubino, Francesca Frontini, Valeria Quochi

Istituto di Linguistica Computazionale “A. Zampolli”
Consiglio Nazionale delle Ricerche
Via Moruzzi, 1 - Pisa, Italy
name.surname@ilc.cnr.it

Abstract

The present paper tackles the issue of PoS tag conversion within the framework of a distributed web service platform for the automatic creation of language resources. PoS tagging is now considered a “solved problem”; yet, because of the differences in the tagsets, interchange of the various PoS taggers available is still hampered. In this paper we describe the implementation of a PoS-tagged-corpus converter, which is needed for chaining together in a workflow the FreeLing PoS tagger for Italian and the DESR dependency parser, given that these two tools have been developed independently. The conversion problems experienced during the implementation, related to the properties of the different tagsets and of tagset conversion in general, are discussed together with the solutions adopted. Finally, the converter is evaluated by assessing the impact of conversion on the performance of the dependency parser by comparing with the outcome of the native pipeline. From this we learn that in most cases parsing errors are due to actual tagging errors, and not to conversion itself. Besides, information on accuracy loss is an important feature in a distributed environment of (NLP) services, where users need to decide which services best suit their needs.

Keywords: PoS tag conversion, interoperability, NLP pipelines

1. Introduction

The possibility to reuse and interoperate language resources and technologies has become a hot theme in current Language Resource Technology (LRT) research, in particular in areas related to the set-up of infrastructures and distributed platforms for sharing or automatically producing Language Resources (LRs) on demand, e.g. META-SHARE (Federmann et al., 2011), the UIMA¹ based U-Compare platform (Kano et al., 2011), GATE (Cunningham et al., 2002), WebLicht (Hinrichs et al., 2010).

Deploying NLP and tools as web services opens up new possibilities not only within the research community, but also for commercial players. However, before such scenario of distributed sharing and use of tools and resources becomes a fully operational reality, there are still a number of research and technical issues to be solved.

Mostly these are related to the issue of interoperability, as in fact most of the tools and resources available are independent: developed by different groups, in different projects, for different purposes. Thus, chaining tools that were not built to work together in pipeline is still a non-trivial task.

Such an attempt is made in the EU-FP7 PANACEA project² whose main objective is to develop a Platform of ‘interoperable services’ to be used like a factory that allows to dramatically cut the costs for the production of language resources.

This paper describes the creation of a morphosyntactic pipeline composed by chaining together the FREELING PoS tagger³ (Padró et al., 2010) and the DESR dependency

parser for Italian (Attardi, 2006; Attardi and Ciaramita, 2007), which were not originally meant to work together and thus have different format and tagset requirements. Instead of modifying the tools, which may not even be always possible in a platform like PANACEA, for chaining the two tools a converter is developed.

Specifically, the paper will describe some of the conversion problems experienced during the implementation of a converter from the FreeLing morphosyntactic output (whose tagset is claimed to be EAGLES compliant⁴) to the input required by the DESR parser for Italian (which uses the TanI tagset⁵, CoNNL format), and will assess the impact of such conversion on the performance of the syntactic parser.

This is no empty exercise, but a necessary step within a more complex work flow that will allow a user to fully automatically acquire a lexicon of subcategorization frames for Italian. Also, because of licensing and IPR issues, we had to look for open source or freely available tools to deploy within the platform.

2. The PANACEA Platform

PANACEA is an EU-FP7 funded project with the main objective of building a platform that serves as a factory of language resources. The availability of LRs is still a major bottleneck for most Language Technology applications, e.g. Machine Translation. PANACEA therefore automates the stages involved in the acquisition and production of language resources, thus helping to cut the costs and time for

splitter, tokeniser, PoS tagger and morphological analyser modules, as these are the basic type of information generally needed for various more sophisticated LT tasks.

⁴<http://nlp.lsi.upc.edu/freeling/index.php>

⁵<http://medialab.di.unipi.it/wiki/>

¹<http://uima.apache.org>

²<http://www.panacea-lr.eu>

³FreeLing is actually a library of NLP functionalities. What we refer to here as PoS tagger, is in fact the output of the sentence

their production. Technically, the goal is to develop a platform of interoperable web services. One of the major issues that hamper interoperability is not surprisingly the different input/outputs that the various tools require. Based on a workflow manager, the PANACEA platform allows the user to combine different LR processors, deployed as web services that may be distributed on different servers and can be used by different users from various locations. Web Service Providers (WSPs) can be institutions (universities, companies, etc.) who offer services for use by a community. The platform is based on a set of Bioinformatics technologies developed by myGrid⁶ team within the scope of e-Science: Soaplab(Senger et al., 2003), used to deploy WSs, and Taverna(Missier et al., 2010), used for designing and running workflows. Many different tools have already been deployed and integrated as WSs by some WSPs: from Python tools to UIMA components (Prokopidis et al., 2011). An advantage of using workflows is that the users do not need to install the tools nor to have deep knowledge of the technical aspects involved in all the technology they need to perform a given complex task. The user instead can create a workflow focusing on high level functionalities. In the specific case of this work, we need to chain together a PoS tagger and a Dependency Parser for Italian in order to have input data for a subcategorisation frame acquisition tool (see the Taverna workflow design in Fig.1 below), yet changing or re-training already existing tools in such a scenario is likely not possible, without changing the existing tools themselves .

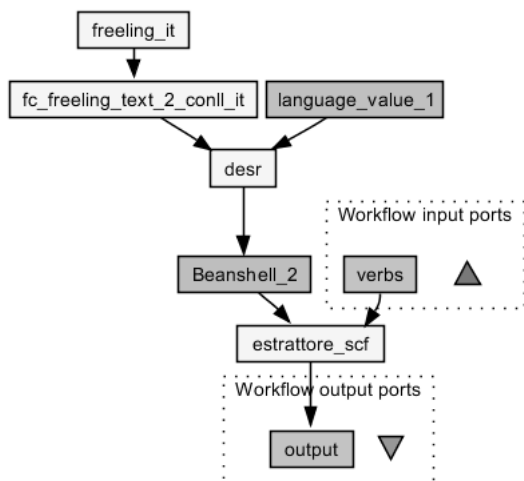


Figure 1: The Taverna Workflow

Due to practical reasons (mainly related to availability and openness), the Italian morphosyntactic pipeline for the PANACEA platform prototype was created by integrating the Freeling PoS tagger and morphological analyser and the DESR dependency parser. In the following sections we will therefore discuss issues and problems tackled in the development of the conversion service to be deployed for chaining these two tools.

⁶<http://www.mygrid.org.uk>

3. PoS Tagging and Interoperability

Automatic PoS tagging *per se* is considered to be a “solved problem” in NLP, as demonstrated by the high level of accuracy of esp. statistical PoS taggers.

Yet, because of the differences in the tagsets used, wider use of the various PoS taggers available (free or open and already trained) is still hampered. With respect to interoperability, thus PoS tagging still poses a number of problems. Disregarding theoretical issues related to lexical category attribution, which are not in focus here⁷, one of the most crucial issues from the practical point of view is that there is not yet a widely agreed upon common set of tags for all languages.

From the practical point of view PoS tags are often a combination of information about the lexical category of the lemmas and the morphosyntactic features of the tokens. While the tagset for main lexical categories is finite and relatively small (for instance PoS tags may distinguish between finite and non finite verbs, between main verbs, auxiliaries and modals, between numerals and other adjectives), adding morphosyntactic features results in an explosion of possible tags. For Standard Average European languages this means adding information about at least gender, number, case, tense, aspect, degree. The intersection of all these multiple dimensions of analysis produces tagsets which may be as large as 274 (Leech and Wilson, 1999). Fortunately, in many cases tags are compositional, so that mapping and conversion can be proceduralised (thanks to the EAGLES initiative).

4. Converting from the Freeling to Tanl

Freeling PoS tags, in line with the EAGLES specifications (Monachini, 1996) have the form of a single positional code (e.g. VMIP3S0), where the first position of the code is always the main category (verb, adjective,...), and the other positions have values that depend on and may specify the sub-category and various morphological traits (e.g. mood and tense for verbs, gender for nouns and adjectives, person, number, etc.).

The Tanl tagset is also compliant to the EAGLES guidelines, but does not make use of positional codes. In fact, it represents the same number and type of information, but in a tab separated form (e.g. V VA num = s|per = 3|mod = i|ten = p). Thus, while format conversion is simple, translation of one tagset into the other requires deeper knowledge of the tag semantics⁸.

When mapping a source tagset into a target tagset four cases can take place; following Teufel (1995):

1. **1:1** the source tag corresponds exactly to the target tag and only renaming is required;
2. **n:1** the source tagset makes finer distinctions than the target tagset, thus more than one source tags are collapsed into one target tag;

⁷see (Aarts, 2007) for an overview

⁸We estimate a total effort of 1 week person for developing the converter, most of which is spent to understand the semantics esp. of the Freeling tags and to study the cases of not direct correspondence.

3. **1:n** the target tagset makes finer distinctions than the source tagset; in this case direct conversion output will result in information loss.

4. **n:m** the class of tokens that is selected by the source tag is only partly overlapping with a target tag

In our exercise we encountered all 4 cases.

In the first two cases (which account for most of the categories: e.g. verbs, nouns, conjunctions, prepositions, etc.) conversion is relatively straightforward and direct mappings can be defined. Also, information loss in the second case does not pose problems in this case as it is not expected to have an impact on parsing performance, given that the parser does not expect that kind of information at all.

An example of **n:1** mapping are attributive adjectives. Freeling distinguishes them by degree (zero and comparative), while Tanl does not, so they are both converted into the Tanl tag A.

The two latter cases, instead, are more complex, since they require to generate categorial distinctions that are not present in the original tagset. Cases of **1:n** mapping occur in the case of pronouns, adverbs, and articles: Tanl distinguishes between interrogatives pronouns and relative pronouns, negations and adverbs, definite and indefinite articles, predeterminers and indefinite pronouns, while Freeling doesn't make such distinctions and in some cases assign words to a different category than Tanl. For instance, Freeling treats articles as determiners without distinguishing between definite and indefinite, while Tanl has two disjoint categories: one for articles and one for determiners.

Cases of **n:m** mapping involve the classification of several types of numerals, of the light verbs "stare" and "fare" (distinguished from other verbs in Freeling but not in Tanl), and of punctuation tags, which are treated differently in the two tagsets.

In all these cases, heuristics are required in order to obtain a more accurate conversion.

The present converter, implemented in java and deployed as a web service within the PANACEA platform⁹, thus makes use of heuristics that read local information from the Freeling annotated output corpus, i.e. the lemma and the morphological information of the target token, in order to try and re-map each token onto the most appropriate target tag. We chose, however, not to implement heuristics that look at the larger context (preceding or following words) since this would be more similar to a (local) re-tagging then to a conversion, which is usually done token by token, and would be more costly.

In the next section the performance of the Freeling-to-dependency parsing workflow is evaluated, and in 5.1. the conversion heuristics are discussed in light of the impact they have on parsing.

5. Evaluation and Discussion

In this section we report on an evaluation exercise allowing us to assess the impact of conversion¹⁰. To this

⁹<http://registry.elda.org/services/213>

¹⁰Retro-conversion, suggested by one of the reviewers, was not considered a meaningful evaluation in this case as it would basi-

end we evaluate the performance of our *Freeling + converter+DESR* work flow (*Freeling+DESR* henceforth) against the CONNL 2007 gold standard, which is annotated with Tanl, both at the morphological and at dependency levels¹¹, and compare it with the native Tanl pipeline (Tanl+DESR)¹²

This allows us to check for and analyse the differences and errors both in tagging and in parsing, as it may be the case that a tagging error does not produce any effect on parsing, and thus can be disregarded (at least for the present purposes). In order to assess the actual impact of the converter, we also (manually) created a "Freeling PoS tagged gold standard" of the same test data¹³, passed it to the converter and then parsed with DESR (*FreelingGold+DESR* henceforth) and evaluated the same way.

The overall results are reported in Table1¹⁴, while Table2 and Table3 show a detailed report of the breakdown of the error rate over the categories.

From these results, we learn that the overall impact on dependency parsing performance of conversion is around 3%, while about 10% information loss is due to Freeling tagging errors or differences.

From the qualitative perspective, the most common occurring problem is that the parser fails to recognize the root, i.e. the main verb of the sentence, which then affects and compromises the correct parsing of the whole sentence (98 times, which accounts for 146 mismatches). In many cases this is due to mis-tagging of the verb as a noun, such as in 5.1. where the verb *corso* "run(pass.part)" is mis-tagged (and lemmatised) as the noun *corso* "alley, course".

- (1) Paolo è corso alla caserma dei carabinieri
'Paolo has run to the police station'

From the conversion point of view we expected to have more severe problems with the $1 : n$ and $n : m$ cases. At a closer inspection, we find out that in fact only part of the mis-parsing is due to conversion, such as the ones that involve light verbs (in fact all cases of errors in the verb category are related to this). The others are more likely to be generated by the different behaviour of the two taggers, such as the ones involving pronouns, where the differences in assigning morphological features to the pronouns may

cally require the implementation of a second converter from Tanl to Freeling which would present similar problematic issues. In particular simple $n:1$ conversion cases in one direction would become more complex cases ($n:1$) in the other, for which completely new heuristics would be needed.

¹¹<http://poesix1.ilc.cnr.it/evalita2009/>

¹²<http://medialab.di.unipi.it/wiki/Tanl>. We thank Dr. Dell'Orletta for parsing the data for us and providing the output results.

¹³I.e. by correcting Freeling mis-tagging and lemmatisation. In fact, Freeling has a peculiar lemmatisation approach, which causes some mis-parses - mostly mislabelling of certain relations, such as the indirect object one.

¹⁴Please notice that these results should not be taken as realistic evaluation of the performance on parsing both because it is likely that the DESR system we use included the gold standard as part of the training set, which explains its very high accuracy (Dell'Orletta 2011, personal communication).

	Freeling+DESR %	Tanl+DESR %	FreelingGold+DESR %	$\Delta 1\%$	$\Delta 2\%$
Labeled attachment score	84.60	99.40	94.95	14.80	4.45
Unlabeled attachment score	88.21	99.58	97.08	11.37	2.50
Label accuracy score	89.57	99.68	96.41	10.11	3.27

Table 1: Comparison of overall performances of the Freeling+DESR, the Tanl+DESR, and the Freeling-gold+DESR pipelines against the Evalita2009 test set. In both cases Freeling refers to the data analysed with Freeling and then converted with the converter discussed here. $\Delta 1$ refers to the difference between Freeling+DESR and Tanl+DESR, while $\Delta 2$ refers to the difference between FreelingGold+DESR and Tanl+DESR, which mostly represents loss due to conversion.

Error Rate	words	head err	%	dep err	%	both wrong	%
total	5005	590	12%	522	10%	341	7%
S	1203	128	11%	131	11%	96	8%
F	783	120	15%	39	5%	22	3%
V	744	134	18%	124	17%	105	14%
E	716	64	9%	78	11%	29	4%
R	402	10	2%	2	0%	1	0%
A	325	26	8%	22	7%	20	6%
P	242	23	10%	59	24%	19	8%
B	241	31	13%	24	10%	19	8%
C	191	39	20%	33	17%	23	12%
N	93	8	9%	6	6%	5	5%
D	43	1	2%	0	0%	0	0%
SA	11	2	18%	3	27%	1	9%
I	7	4	57%	1	14%	1	14%
T	3	0	0%	0	0%	0	0%
X	1	0	0%	0	0%	0	0%

Table 2: The overall error rate of *Freeling + DESR* and its distribution over coarse grain PoS tags.

affect the outcome of attachment and labelling. Indeed, there is little evidence that, for $1 : n$ and $n : m$ cases, the conversion *per se* causes loss of accuracy, as it seems that the DESR still manages to assign correct parse.

In the following section we will discuss in more details the heuristics applied by the converter and their effect on parsing results.

5.1. Issues on Mapping Heuristics

Heuristics can be either lossless, i.e. when they perfectly reproduce the native Tanl tagging, or lossy, i.e. when the resulting tagging lacks some information that would have been present in the case of native tagging with Tanl. In the latter case we can also distinguish between harmless loss, when the loss of information has no impact on the subsequent parsing, and damaging loss, when parsing is affected. In this section we discuss problematic mapping issues and the chosen solutions to them, with a special attention to the impact that these solutions have on parsing with DESR.

From this assessment, we observe that the only case of lossless heuristics is the one allowing us to split Freeling articles into Tanl definite and indefinite articles according to the lemma (an example of $1:n$ conversion). Manual inspection shows no mistagging as a consequence of this heuristics.

Other cases, especially those involving $n:m$ mapping, are more complex and heuristics can only approximate the desired target tagging. Finally, there are cases for which no

heuristics has been even attempted because they would be too difficult or too ineffective. When using no heuristics a distinction that exists in the target tagset will be absent in the converted output; when using “imperfect” ones some errors may be introduced. In both cases this may mislead the parser and thus strongly affect parsing performance. Our manual inspection however shows that not all such cases are damaging, since the category that the converter assigns in those complex cases, while constituting an incorrect or incomplete tagging in Tanl, often generates the same parse. In the following some interesting cases are discussed.

An example of missing categorial distinction is given by predeterminers. Tanl has a special tag for predeterminers, such as *tutto* ‘all’ or *entrambi* ‘both’, while Freeling analyses them in the same way as indefinite pronouns. The status of predeterminer in Italian is not widely accepted and used in all grammatical formalisms/theories, furthermore it is highly context sensitive: most predeterminers in Italian (usually quantifiers) can also be used as nouns or pronouns. Therefore, lemma based heuristics is hardly applicable (without looking at the wider context) and will often end up in mistagging. Our choice was then to treat these cases as a $1:1$ mapping and to convert predeterminers always as Tanl indefinite pronouns. The predeterminer tag will not thus be found in a Freeling-to-Tanl conversion and whatever rule the parser has (learned) involving predeterminers, will not be applied. Interestingly enough, when looking at

Error Rate	words	head err	%	dep err	%	both wrong	%
total	1921	56	3%	69	4%	28	1%
S	471	10	2%	10	2%	4	1%
F	319	14	4%	2	1%	2	1%
V	301	15	5%	18	6%	14	5%
E	281	7	2%	17	6%	1	0%
R	138	0	0%	0	0%	0	0%
A	120	1	1%	1	1%	1	1%
P	88	4	5%	15	17%	3	3%
B	85	3	4%	1	1%	1	1%
C	61	1	2%	4	7%	1	2%
N	31	1	3%	1	3%	1	3%
D	18	0	0%	0	0%	0	0%
I	5	0	0%	0	0%	0	0%
T	2	0	0%	0	0%	0	0%
SA	1	0	0%	0	0%	0	0%

Table 3: The overall error rate of *FreelingGold* + *DESR* and its distribution over coarse-grain PoS tags.

the parsing output from the converted PoS tagged test set, we find out that this information loss or mistagging does not result in mis-parsing. For example, in the sentence:

- (2) Abbiamo offerto aiuto a tutti i serbi di Krajina
‘We offered help to all Serbians from Krajina’

the parser behaves in the same way when *tutti* is recognized as a predeterminer (T) as when it is tagged as an indefinite determiner (DI) following the conversion. In both cases “a tutti i serbi” is parsed as a prepositional phrase introduced by “a”, with “serbi” as the head and “tutti” as a modifier.

A similar case happens with the treatment of numbers: *Freeling* has a special code for numbers, that can be translated into the Tanl code for cardinal numbers. Tanl, however, has also a special code for ordinal numbers, which *Freeling* does not recognize treating them as either nouns or attributive adjectives. As in the previous case, no heuristics is attempted, and therefore ordinals will appear in the converted Tanl tagging as either an adjectives or nouns, with the tag for ordinals never occurring. Also in this case conversion does not affect parsing, and ordinal numbers are correctly treated as heads (when they are nouns) or modifiers (when adjectives) accordingly.

Another case for which we found no impact on the final parse is the interrogative pronoun heuristics. *Freeling* in fact does not distinguish when a pronoun is relative and when it is interrogative, and it always assigns the tag for interrogative pronouns. Tanl instead has such distinction and apparently *DESR* has been trained to recognise it.

In order to distinguish between relative and interrogative pronouns (as in (3) and (4)) the converter checks for wh-question lemmas (“chi, cosa, quanto”¹⁵).

- (3) L’immagine che hai visto
‘The picture that you saw’
- (4) Gli ho chiesto chi è venuto
‘I asked him who came’

This rule works for the majority of cases, but introduces errors in some less frequent uses of *chi* as a relative pronoun. E.g.:

- (5) Puoi mandare chi vuoi alla conferenza.
‘You may send whoever you want to the conference’

In all inspected cases however this mis-labelling did not impact on the capacity of the parser to recognize the function (subject or object) of the pronoun.

Other conversion issues instead do result in mis-parsing. This is the case for example with the verbs *stare* (stay) and *fare* (do). These verbs have two functions in Italian; one as main verbs and one as light verbs in (various) periphrastic constructions such as in (6) and (7).

- (6) Giovanni sta facendo la spesa.
‘Giovanni is doing shopping (progressive construction)’
- (7) Maria fa scappare il gatto.
‘Maria makes the cat run away (causative construction)’

Freeling does not disambiguate and assigns a specific code (VS and VF respectively) to all instances of the verbs. Tanl assigns V (main verb) or VA (auxiliary) accordingly to the function. It is very hard to define a rule that distinguishes them on the basis of lemma and morphology alone (without checking the context), so we chose to collapse all VSs and VFs onto Tanl main verbs. In this case all instances of *stare* and *fare* with support function will be mistagged, with the consequence that the parser fails to identify the main verb as in:

- (8) *Marco sta cercando di dormire.*
subj !main-V !V-mod PP-prep PP-head
subj aux main-V PP-prep PP-head
‘Marco is trying to sleep.’

¹⁵These are generalisations based on most frequent cases.

Another case where a heuristics would have been too complex is abbreviations (e.g. *km*, *dott.*); they receive a special tag in Tanl (NA), while in Freeling they are treated as simple nouns. Labelling abbreviations as if they were normal nouns causes mis-parses in some cases, especially in the case of measures. In *circa 130 km* ('about 130 km') the parser behaves differently when *km* receives S (noun) or SA (abbreviated noun): in the latter case it correctly allows the parser to recognize the abbreviation as the head.

Some other cases are even more complex and cannot be considered as simple conversion issues. We noticed that even in the categories for which the Freeling and Tanl tagsets have a good correspondence in general, in some marginal cases these categories are applied in different ways. This is an intrinsic limitation due to tagging strategies and can hardly be overcome without basically changing the behaviour of the tagger during the conversion. This may cause problems since each parser is designed for - or trained on - the behaviour of a given tagger (as in this specific case DESR was trained on a Tanl gold standard).

Here again some outcomes are harmless and do not affect the parser, as it happens in the case of the clitic "ci" in the frequent verb *esserci* ("c'è", there is), which is labelled as a pronoun in Tanl and as an adverb in Freeling (grammatically both could be considered correct labelling, since "ci" is a locative pronoun that stands for an adverb "there"). This notwithstanding, the parser treats both labelling outcomes in the same way.

Other cases are more problematic; for instance in

- (9) *Qualcuno si fa male*
 Someone REFL does harm
 'someone gets hurt'

the Tanl gold standard tags *male* as an adverb, while the Freeling labels it as a noun. These two tags may correspond to two interpretations of the function of the word. On the basis of the different labels the tagger considers it in one case a predicative complement of the verb and in the other case the object.

Another interesting case is the morphological analysis of pronouns. The genitive clitic pronoun "ne", as in

- (10) *Se ne occupano i preti*
 REFL of this concern the priests
 'The priests are taking care of this'

receives a third person morphological feature tagging in Freeling while it remains person neutral in TANL. The former tagging causes the parser to attach the clitic to the verb as subject (instead of complement), although this lemma can never occur in such function in Italian. This is clearly due to the fact that DESR was trained to recognize person agreement between a pronoun and a verb as a clue of subjectivity.

Finally and interestingly enough, the most severe parsing errors are introduced by categorial distinctions that may seem "minor" at first observation, such as the ones affecting punctuation. Freeling has a code for each punctuation mark and basically avoids all attempts to recognize the function

of each punctuation. Tanl instead distinguishes among balanced punctuation, clause boundary punctuation (such as semicolon), comma and sentence boundary punctuation. Punctuation also seems to be an important piece of information used by the dependency parser to establish the dependencies (to decide attachment). Given the importance of punctuation, heuristics were used to map each symbol onto the most likely Tanl function, but error is introduced. For instance the Tanl gold standard distinguishes between the uses of "-" as balanced punctuation, such as in:

- (11) *La mia sensazione - dice - è che la nostra regione sia gestita come un grande comune.*
 'My sensation - he explains - is that our region is managed as a big municipality'

and as clause final when it is single. In the given example the parser will consider the portion of text before and after "- dice -" as one sentence; thus correctly identifying the subject relation between "la mia sensazione" and the verb *essere*; mis-parsing occurs instead when the first and the second "-" are marked as clause boundaries, producing three different clauses.

The latter observations lead us to reconsider the priorities in conversion heuristics; when writing a converter a linguist may tend to concentrate on very subtle distinctions for the main categories and neglect other distinctions for the minor ones. Still from the parser's perspective the functional role of the different types of pronouns does not vary (especially from the point of view of the attachment, they are mostly heads); on the contrary the mis-interpretation of a punctuation mark often has large effects on phrase boundary detection and main verb identification, an effect which sometimes propagates over several lines.

6. Conclusion and Future Work

In this paper we describe and discuss a problematic issues related to PoS tag(set) conversion and evaluated a conversion software that allowed us to chain together the Freeling Italian morphosyntactic tagger and the DESR dependency parser, which are both open source, developed independently and use different data formats and tagsets. Evaluation shows that, notwithstanding some information loss due to conversion (esp. for 1:n, n:n tag correspondences), the conversion itself has little impact on the overall performance, while, not surprisingly, actual tagging errors do. However, overall performance is still relatively high, which can be acceptable depending on the application purposes of the workflow outcome. This in turn is encouraging, as it suggests that tool interchange through distributed platforms and/or infrastructures is indeed a viable approach without necessarily having to modify/adapt the tools themselves, albeit accepting some performance loss. Of course, in the specific case of this paper, training the dependency parser on the Freeling annotation or viceversa training Freeling on a Tanl annotation, or even using a different parser trained on Tanl annotation would be better solutions. However this may not always be possible/desirable for users, as creating training sets is costly and time consuming, while ready to use tools may be a good compromise choice. The present

exercise also shows that an assessment of the impact of one tool on the performances of another, as well as the impact of conversion, constitutes useful information for potential users of such a platform, as this shall help them deciding what tools to choose (if a choice is available) for building their workflows.

As a future development the converter could be implemented to natively work on GrAF (Ide and Suderman, 2007), which has been chosen as the standard corpus format for PANACEA and attempts to generalise the converter could be done as well.

On the evaluation side, we intend to perform other experiments so as to study more accurately conversion issues and the impact on the overall dependency parsing results of: 1—different PoS taggers; 2—different conversions strategies.

7. Acknowledgements

This work has been funded within the EU funded FP7-ICT project PANACEA: Platform for Automatic Normalized Annotation and Cost-Effective Acquisition of Language Resources for Human Language Technologies (GA:248064).

8. References

- Bas Aarts. 2007. *Syntactic Gradience: The Nature of Grammatical Indeterminacy*. Oxford Linguistics. Oxford University Press.
- Giuseppe Attardi and Massimiliano Ciaramita. 2007. Tree revision learning for dependency parsing. In Candace L. Sidner, Tanja Schultz, Matthew Stone, and ChengXiang Zhai, editors, *HLT-NAACL*, pages 388–395. The Association for Computational Linguistics.
- Giuseppe Attardi. 2006. Experiments with a multilanguage non-projective dependency parser. In *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X)*, pages 166–170, New York City, June. Association for Computational Linguistics.
- Hamish Cunningham, Diana Maynard, Kalina Bontcheva, and Valentin Tablan. 2002. GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02)*.
- Christian Federmann, Byron Georgantopoulos, Riccardo del Gratta, Bernardo Magnini, Dimitris Mavroei-dis, Stelios Piperidis, and Manuela Speranza. 2011. Meta-share functional and technical specification (deliverable d7.1). Technical report, METANET, January. Dissemination Level: Restricted.
- Erhard W. Hinrichs, Marie Hinrichs, and Thomas Zastrow. 2010. Weblicht: Web-based lrt services for german. In *ACL (System Demonstrations)*, pages 25–29. The Association for Computer Linguistics.
- Nancy Ide and Keith Suderman. 2007. GrAF: A graph-based format for linguistic annotations. In *Proceedings of the Linguistic Annotation Workshop*, pages 1–8, Prague, Czech Republic, June. Association for Computational Linguistics.
- Yoshinobu Kano, Makoto Miwa, K. Bretonnel Cohen, Lawrence Hunter, Sophia Ananiadou, and Jun ichi Tsujii. 2011. U-compare: A modular nlp workflow construction and evaluation system. *IBM Journal of Research and Development*, 55(3):11.
- Geoffrey Leech and Andrew Wilson. 1999. Standards for tagsets. In H. van Halteren, editor, *Syntactic Wordclass Tagging*, number 9 in Text, speech, and language technology, pages 55–80. Kluwer Academic Publishers, Dordrecht.
- Paolo Missier, Stian Soiland-Reyes, Stuart Owen, Wei Tan, Aleksandra Nenadic, Ian Dunlop, Alan Williams, Thomas Oinn, and Carole Goble. 2010. Taverna, reloaded. In M. Gertz, T. Hey, and B. Ludascher, editors, *SSDBM 2010*, Heidelberg, Germany, June.
- Monica Monachini. 1996. Elm-it: Eagles specifications for italian morphosyntax lexicon specification and classification guidelines. Technical report, Paris, France, France.
- Lluís Padró, Miquel Collado, Samuel Reese, Marina Lloberes, and Irene Castellàn. 2010. Freeling 2.1: Five years of open-source language processing tools. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors, *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may. European Language Resources Association (ELRA).
- Prokopis Prokopidis, Byron Geogantopoulos, and Haris Papageorgiou. 2011. A suite of nlp tools for greek. In *Proceedings of the 10th International Conference of Greek Linguistics*, Komotini, Greece.
- Martin Senger, Peter Rice, and Thomas Oinn. 2003. Soaplab - a unified sesame door to analysis tools. In *All Hands Meeting*, September.
- Simone Teufel. 1995. A support tool for tagset mapping. In *Proceedings of SIGDAT 1995. Workshop in cooperation with EACL 95, Dublin*.