Dependency Relations for Negation in Spoken French

NADINE GLAS

MA Thesis
Erasmus Mundus Language and Communication Technologies

June 2011

Rijksuniversiteit Groningen
Université Nancy II

Supervisors:
Corinna Anderson, Christophe Cerisara, Gertjan Van Noord

Contents

Introduction

In 2010 Cerisara, Gardent and Anderson have started the construction of a dependency treebank for French radio broadcast (Cerisara, Gardent, Anderson, 2010). The proposed dependency relations are based on an annotation guide. However, the annotation guide is not complete yet. One of the issues left for a more thorough consideration is negation. Here we will have a closer look at negation in spoken French in order to propose adequate dependency relations for negative items. An attempt to automatically implement the proposed annotations shall be demonstrated as well. Adding the annotation scheme for negation to the general annotation guide of the radio broadcast Treebank will increase the phenomena covered by the guide and should therefore improve the consistency of annotations.

During the development of the annotation scheme for negation two different ways of reasoning are required. On the one hand, linguistic characteristics and notational issues about dependency grammar give deductive clues about what, according to these formalisms, would be a justified annotation. But at the same time, proposing annotations is a prescriptive and thus inductive way of dealing with language. Annotations do not only have to deal with linguistic characteristics and notational issues but also with other dependencies in the annotation scheme. It has to be made sure that dependencies do not conflict with each other and that the annotations are consistent. For the automatic implementation of annotations even more issues will need to be dealt with.

This project aims at proposing annotations for negation that are plausible treatments of negative items, from all different frameworks. To define the areas that will be worked with first some theoretical background shall be provided. In section 2 all thinkable constructions of the negative items under consideration (*pas, plus* and *jamais*) will be discussed and corresponding annotations shall be proposed. The constructions in which the negative items can occur are collected not only from the corpus of spoken French, but also from studies of negation by other authors and native speaker judgements. In this way a possible increase of the spoken corpus will be anticipated. Section 3 proposes a way to automatically implement the annotations. After this an evaluation will follow about the proposed annotations and in section 5 the automatic implementations of the annotations will be evaluated. In the last section the project will be concluded.

1   Theoretical Background

In order to develop an annotation scheme for French negation in dependency grammar background knowledge is needed regarding the corpus to which the annotation scheme should apply, dependency grammar and of course French negation. Here a short introduction is given to all of these areas.

1.1   Corpus

The corpus for which an annotation scheme regarding negation shall be developed is the French dependency treebank developed by Cerisara, Gardent and Anderson (2010). This is the first treebank for transcribed common French speech. At this moment it includes 50,000 words representing one hour of speech.

The data from the Treebank comes form the development set of the ES-TER corpus of manual transcripts for French radio news. These transcripts do not describe all the features of speech like incomplete words, laughs, and noises for example, but do include hesitations (*euh*). The original transcriptions were grouped into prosodic segments but the sentence boundaries could be adapted from annotators of the Treebank if there were incomplete constituents.

Part for part, the corpus is provided with part of speech tags automatically assigned by the TreeTagger for French. Then, one part of the corpus was automatically parsed with the Malt Parser trained on another 20,000 words corpus. Afterwards, linguistic students manually segmented the utterances and checked the dependency labels by means of an annotation guide. The malt parser was then retrained on this part and automatically annotated new parts. Validation of the annotations was performed by discussions between annotators and an expert linguist where difficult cases were solved. Before elaborating on the details about the dependency relations in the corpus we shall now first explain the idea of dependency grammar in general and then come back to the annotations in the corpus.

1.2   Dependency grammar

Dependency grammar is developed by the French linguist Lucièn Tesnière in 1959. The goal of this grammar is to represent syntactic *relations* outside the linear appearances of language. A so-called *dependency relation* occurs between two items in a sentence where one of them is a dependent of the other. For example, a noun is in a dependency relation with a verb when there is syntactic relation between them, like a subject relation where the noun is the subject of the verb. The element that syntactically depends on the other element is called the *dependent* and the element of which it depends is called the *governor*. In the mentioned example the noun would

be the dependent and the verb the governor. While Tesnière proposed to represent the dependency relations by placing one below the other, dependency relations are nowadays often graphically represented by arrows which point from the dependent to the governor. They are labelled with the type of the dependency relation that occurs between the connected items. The dependency relation between a subject a verb looks for example as 1a. The base of the arrows can be recognized by a little black triangle indicating that an element functions as dependent while the other side of the line points to the dependent's governor.
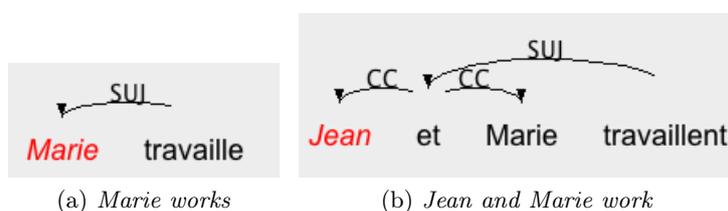

(a) *Marie works*   (b) *Jean and Marie work*

Figure 1: Examples of dependency relations

There are different kind of syntactic relations that can occur between items. Another example can be seen in 2 where there are two dependents of the conjunction *et* attached to their governor with a *CC* labeled dependency relation. How many and what kind of dependency relations are distinguished is defined for every project separately.

Every item can only have one dependency relation to its governor but can have multiple dependents, (as can be seen in 2). When a sentence is completely annotated, every item has a dependency relation to its governor. All items are in this way directly or through other governors dependents of the verb. This makes the verb the governor of all elements in the sentence and the highest element in the tree. This is unlike many other syntactic formalisms such as phrase structure grammar, where not the verb but the subject is the highest element in the tree.

Phrase structure grammar is a syntactic representation that is more widely known and used than dependency grammar. The main difference between phrase structures and dependency relations is the fact that phrase structure grammar is based on *constituents* instead of *relations*. Phrase structure grammar respects the linear realization of the items since constituents are connected in a way that corresponds with the order of the items in the sentence. From the bottom to the top of a syntactic tree for phrase structures, constituents are connected to each other and form constituents of a higher order. Looking top-bottom, the tree reveals the structure of an expression by cutting the complete sentence in constituents which in their turn are cut in smaller constituents. Phrase structure trees thus represent what items belong to a certain phrase or category. In contrast to this, de-

3

pendency relation trees reveal the structure of an expression not in terms of constituents but in hierarchical links between the elements (Mel'čuk, 1988). It shows how items are syntactically related without taking into account the linear realization of the sentence. Below the difference in representation among the two formalisms can be seen by the same sentence represented in the two different ways.

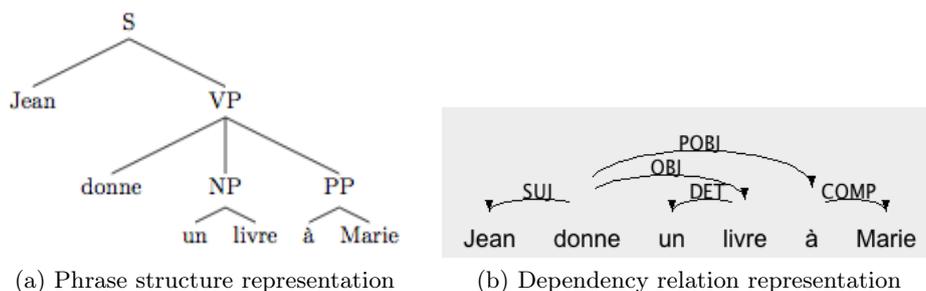(a) Phrase structure representation   (b) Dependency relation representation

Figure 2: *Jean gives a book to Marie*

It can be seen that constituents like NPs (noun phrases) or VPs (verb phrases) are restricted to phrase structure grammar. However, the items inside a constituent are also elements that are connected to each other by means of dependency relations. This means that the representations for both grammars do have similarities in their structures. After all, the syntax is the same, only represented differently. When dealing with dependency grammars we can thus also talk about constituents, not because they are labelled as such, but simply to refer to a certain sequence of words that are related to each other.

The theory Tesnière has developed is much more exhaustive than there is place to describe here. Here only the core of the idea is presented, necessary to understand the essence of the grammar. In the following paragraph it shall be discussed how dependency grammar is exactly used in the corpus for which dependency annotations will be proposed.

1.3 Dependency relations in the Treebank

The dependency Treebank Cerisara, Gardent and Anderson (2010) develop is based on the already existing ESTER corpus, enriched with, amongst others, dependency relations. The principle of these dependency relations is the same as when Tesnière presented it in 1959: There are dependents and governors that structurally connect by means of dependency relations. In this way all elements are either subordinated or coordinated below the governor of the entire sentence, the verb. The dependency relations together structurally represent all elements outside the limits of linear realization.

As mentioned before, the specifics of the implementations of dependency grammar differ for different projects. The annotations added to the ESTER

corpus to construct a French Treebank rely on an annotation scheme defined by previous work of the developers on the Syntex parser and the P7Dep corpus (Cerisara, Gardent, Anderson, 2010). It distinguishes 15 different dependency relations with the following labels: SUJ (subject), OBJ (object), POBJ (prepositional object), ATTS (subject attribute), ATTO (object attribute), MOD (modifier), COMP (complementizer), AUX (auxiliary), DET (determiner), CC (coordination), REF (reflexive pronoun), JUXT (juxtaposition), APPOS (apposition), DUMMY (syntactically governed but semantically empty dependent), DISFL (disfluency). Some of these labels are specifically added to the annotation scheme because their relation occurs frequently in transcribed speech. This is the case for the juxtaposition (to relate constituents which occur in the same prosodic group but are not syntactically related), the apposition (to relate two adjacent constituents denoting the same referent) and disfluencies (to relate hesitations, repairs and missing words). The annotation scheme for negation that will be added to the already existing annotation scheme of the dependency Treebank can thus either work with the already defined labels, or new labels can be created when the existing ones appear not to be enough. Another characteristic of the annotations in the corpus need mentioning is that the graphical representation of the relations is not the same as Tesnière originally proposed. Instead of presenting one element below another the hierarchy of elements is presented with horizontal arrows as demonstrated in figure 1. This representation is used in the software environment *jsafran*, which is specially developed for the dependency annotation of the corpus. It was also used in the project for the development of the annotations of negation. More details about the tool can be found in Cerisara, Gardent, Anderson (2010) and http://www.loria.fr/~cerisara/jsafran/index.html. A way to represent dependency relations in text is the following: dependent = TYPEOFDEPENDENCY (governor). For example the noun *Marie* that has a subject dependency relation to a verb *travaille* (1a) can be represented as: Marie=SUBJ(travaille). This is the way dependency relations shall mainly be represented in this project.

A last comment about the existing annotations of the French dependency corpus regards the already mentioned manner of adding annotations. The corpus was annotated by a combination of training, parsing and manual corrections. First a part of the corpus was automatically parsed by the Malt parser, which was trained on another 20000 words corpus. Human annotators had to check the proposed dependencies and possibly modify some of them. Discussions between the annotators and an expert linguist solved the ambiguous and difficult cases (Cerisara, Gardent and Anderson, 2010). After that, a second part of the corpus was annotated with the parser now trained on the annotations of the first part, and so on, until the entire corpus was annotated. For the annotations of dependency relations for negation, it would be consistent to have the same approach of alternating

between techniques. Due to the short time and number of people available for this project it will however not be possible to make use all of these. Besides, handling negation requires profound study of negation and checking for consistency, which has not been done for the applications of the existing annotations.

1.4 Negation

Besides knowledge about the corpus and its dependency relations a short introduction to French negation in necessary to provide some background about what types of French negation exist and what parts of this area shall be covered in the proposed dependency scheme.

1.4.1 Logical versus semantic negation

First negation can be devided in 'logical' negation and items whose semantics have a negative flavor, such as *heureux/malheureux* (*happy/unhappy*) and *accepter/refuser* (*accept/reject*) (Corblin, 1994, 1996 in Godard, 2004). Here we will focus on logical negation only.

Logical negative expressions in French are different from negation in other Romance languages in that they cannot be recognized on a morphological basis (Godard, 2004). This is because historically negation was expressed by the item *ne*. Then, in Medieval French the item *pas*, meaning step, was added to emphasize the negation (Jones, 1996). For example, *il ne marcha pas* meant *he did not walk a step*. Later *pas* extended its use to all verbs until it lost its meaning of step and achieved the status of a negative particle (Jones, 1996). Now standard negation is expressed by *ne pas* (*not*).

1.4.2 Negative expressions' part of speech categories

Although, negation in French has no strict morphological basis starting with some *n-*, all the possible negations are associated with the verb clitic *ne* (more about this in the following paragraph) and can be recognized by a list of negating words (Godard, 2004): There are negative NPs: *personne* (*nobody*) and *rien* (*nothing*), NPs or determiners: *aucun, nul* (*no, none*), *pas un* (*not one*), negative adverbs: *pas* (*not*), *plus* (*no more*), *jamais* (*never*) and *aucunement, nullement* (*in no way*), *guère* (*not*) and *point* (*not*), the negative preposition *sans* (*without*), and the conjunction *ni* (*neither*). Lastly, there is the pro-S adverb *non*, which can be used as an answer or as an embedded complement sentence (*Paul pensait que non*, 'Paul thought that is was not the case') which is also considered a negation.

Godard distinguishes different kind of negating elements by assigning them part of speech categories but that does not necessarily mean that they act exactly in the same way as other members in this category. Negations are special items within a category. Other authors mention this explicitly:

negative words have only characteristics in common with certain part of speech categories (Rowlett, 1998). We can therefore not derive the behavior of negations from other members of a POS category but can only notice their similarities (or differences).

Because of time constraints we cannot cover all negative expressions in the annotation scheme, we shall select only the most frequent negative item and its most frequent category members, namely the negative adverbs: (*ne*) *pas*, (*ne*) *plus* and (*ne*) *jamais*. Other negative adverbs occur little or never in the corpus and will not be considered in this project (*aucunement, nullement* and *guère*). By explaining the role of *ne* in combination with the negative *pas*, *plus* and *jamais*, we shall explore adverbial negation further and set out the different roles that the negations can fulfill.

1.4.3 Sential, clausal and constituent negation

Ne is a clitic to the verb and indicates the scope of the negation (Godard, 2004). It ensures that the negation has scope over the sentence where *ne* occurs and not higher. Godard (2004) demonstrates this with some examples for the negation *ne personne* (*nobody/not anybody*):

(1) Paul accepte de **ne** recevoir   **personne**
Paul accepts of *ne* to-receive person
Paul agrees not to see anybody

(2) Paul **n'** accepte de recevoir   **personne**
Paul *n'* accepts of to-receive person
Paul does not agree to see anybody

In both cases *personne* occurs at the same position and in the same clause as *ne*. Only the position of the scope marker *ne* differs, which makes the interpretations of the negations different for both sentences. In the first sentence *ne* indicates that the negation has scope over the verb *recevoir*, which leads to the negation of the clause *de recevoir personne*. In the second sentence the scope marker indicates that *accepter* is negated which means that the truth value of the entire sentence *Paul accepte de recevoir personne* is negated.

Adverbial negation also occurs in combination with the scope marker *ne*, although for these negations the position of the adverbs themselves also differ:

(3) Paul accepte de **ne pas** / **plus** / **jamais** recevoir   Marie
Paul accepts of *ne* not / more / never   to-receive Marie
Paul agrees not/not anymore/not ever to see Marie

(4) Paul **n'** accepte **pas** / **plus** / **jamais** de recevoir   **Marie**
Paul *n'* accepts not / more / never   of to-receive Marie
Paul does not/not anymore/not ever agree to see Marie

According to Godard (2004) both of the constructions shown above are called sentential negation and can be recognized by the (possible) presence of *ne*. In order to know exactly about what construction we are talking in this work, we adapt this terminology a little and reserve the term sentential negation for (the possible) *ne* in front of a finite verb (as in 4), and use the term clausal negation whenever (the possible) *ne* precedes an infinitive (as in 3). The second parts of negation *pas*, *plus* or *jamais*, occur usually after the finite verb for sentential negation and before the infinitive in clausal negation. Because both sentential as clausal negation are situated in verb clauses, they both will be called forms of verbal negation.

The reason why we talk about a possibility of *ne* being present instead of its simple presence is because *ne* is often omitted in spoken or colloquial French. However, as long as *ne*'s presence in combination with a negating word would be grammatically possible without changing the meaning of the sentence, there is still sentential or clausal negation performed. Look for instance at the examples (3) and (4) without the scope marker:

(5) Paul accepte de **pas** / **plus** / **jamais** recevoir   Marie
    Paul accepts of not / more / never   to-receive Marie
    Paul agrees not/not anymore/not ever to see Marie

(6) Paul accepte **pas** / **plus** / **jamais** de recevoir   **Marie**
    Paul accepts not / more / never   of to-receive Marie
    Paul does not/not anymore/not ever agree to see Marie

The structure of the sentences make that *ne* can be added at the places indicated by (3) and (4) without changing meaning or grammaticality. So although *ne* is not present, it is still implicated and a marker for sentential and clausal negation. As seen by the above examples, *pas*, *plus* and *jamais* can all express sentential and clausal negation.

Although *ne* can be absent and implied in case of verbal negation, in its absence it is not always clear whether sentential or clausal negation takes place. This ambiguity arises when the requirements for the constructions of both negations have been met. This is the case when the negating element not only precedes an infinitive but also follows an inflected verb. Look for example at sentence (7) (a sentence from the corpus with *ne* removed). The positions where *ne* could be implied are represented between brackets.

(7) Mais il **(ne)** pourrait **(ne) pas** s'   en tenir   au seul maire de
    But it *(ne)* could   *(ne)* not itself in to-keep at only mayor of
    Paris
    Paris
    But it could not stick to the only mayor of Paris

Here, it is not clear at what place the implied *ne* occurs: before the inflected verb (*pourrait*) or directly before *pas* that is preceding the infinitive (*tenir*).

The first case leads to the interpretation that it is not possible that it sticks (to the only mayor of Paris) while in the second case it will be interpreted that it is possible that it does not stick (to the only mayor of Paris). There is no way to solve this ambiguity because both underlying constructions are syntactically equally possible. However, it must be decided what kind of negation takes place in order to propose appropriate dependency relations. Although both types of negations could occur here, it will be assumed that sentential negation is taking place, based on the fact that this is the most frequent negation. In case of ambiguity between sentential and clausal negation caused by the absence of *ne*, dependency relations based on sentential negation will be proposed.

Besides sentential and clausal negation, there is also *constituent negation*. In this type of negation the negative word is not combined with the clitic *ne* (Godard, 2004). Instead, the negating words *pas*, or *non pas* (Godard, 2004) negate an argument of the sentence on their own. Some authors disagree with this distinction and consider clausal negation (where *ne* and *pas* precede an infinitive) as a form of constituent negation as well (Kim and Sag, 2001). We will follow the terminology of Godard and speak of constituent negation as soon as there is no (implied) *ne*. Rowlett gives some simple examples (1998):

(8) Ça va? **Pas** mal.
it goes not bad
How are you? Not bad.

(9) Qui est- ce qui veut un café? **Pas** moi.
who is it who wants a coffee not me
Who wants coffee? Not me.

(10) **Pas** vrai!
not true
Never!

(11) T' as du fric? **Pas** un sou.
you have of money not a coin
Got any money? Not a penny.

It can be seen that *pas* negates the argument that follows in the absence of *ne*. In similar structures *jamais* and *non plus* can also express constituent negation, as can be seen by some examples (checked by a native speaker):

(12) Ça fait mal quand on retire une dent? **Jamais** mal.
it does pain when we remove a tooth never bad
Does it hurt when one a tooth is removed? It never hurts.

(13) Ça fait mal aussi quand on retire une dent? Mal **non plus** .
it does pain also when we remove a tooth bad non more
Does it hurt as well when one a tooth is removed? It hurts neither.

(14) L' a   t on aidé   durant sa traversée du désert? **Jamais** le
he have *t* one helped during his crossing of dessert never   the
moindre soutien.
lesser   support
Did they help him during his crossing of the dessert? Never any
support.

(15) Je ne veux pas un café. Moi **non plus**.
I *ne* want not a coffee me not more
I don't want a coffee. Me neither.

A special circumstance that in case of negation often leads to constituent
negation is ellipsis. Some examples are (Godard, 2004):

(16) Je ne suis jamais allé à Rome, et   a Paris **non** *plus*
I *ne* am never went to Rome and to Paris not more.
I have never been to Rome, and neither to Paris

(17) On lit   Molière, mais **pas** Corneille
we read Molière but not Corneille
We read Molière, but not Corneille

In this way practically every type of element can follow *pas* or precede *non*
plus without the direct presence of a verb.

A last important thing to note about adverbial negation is that not all
elements that express this kind of negation are inherently negative. *Jamais*
and *plus* can also have positive uses when they occur without the scope
marker *ne*:

(18) À jamais (Rowlett, 1998)
to never
For ever

(19) Plus de pain
more of broad
More bread

For *plus* this is always the case; it is always positive without the (possi-
ble) presence of *ne*. This is the reason why *plus* cannot express constituent
negation.

It can be concluded that while the adverbs *pas*, *plus* and *jamais* can
all express sentential negation and constituent negation, *pas*, *jamais* and
non plus can (also) express constituent negation. In the following chapters
annotation schemes shall be developed for the all the negative uses of *pas*,
plus and *jamais*.

2 Proposal Annotation Scheme

In the following sections we respectively propose annotations for all the negative uses of *pas*, *plus* and *jamais*. We shall do this based on both linguistic features and notational issues regarding dependency grammar. Rules to automatically implement the proposed annotations will be presented in section 3.

2.1 (Ne) Pas

In the background section we have seen that *pas* is capable of expressing sentential, clausal and argumental *negation*. For each of these situations we now look at the roles *ne* and *pas* fulfill and propose corresponding dependencies.

2.1.1 Sentential negation

Sentential negation expressed by *ne pas* (not) consists of *ne* before the inflected verb and *pas* that follows the inflected verb, forming sentences like:

(20) Nous **ne** pouvons **pas** nous permettre de perdre Paris
We *ne* can not us to-allow of to-lose Paris
We cannot allow ourselves to lose Paris

(21) Le suspens se poursuit donc pour les 2 Français qui **n'** ont
The suspense itself continues so for the 2 French who *n'* have
pas terminé leurs rencontres
not ended their meetings
So the suspense continues for the 2 French who have not completed their meetings

Previous studies have demonstrated that *ne* is a clitic to the verb and a scope marker of negation (Godard, 2004). As mentioned in the theoretical background, the role of a scope marker entails where the scope of the negation lies. The scope indicates in what part of the sentence the negation is interpreted. A negative word is interpreted negatively in the same clause as *ne*, only if it occurs in the same clause as the (an implied) *ne*. For example, *pas* in (20) is negatively interpreted in the main clause while in (21) the negation takes place in the relative clause. Both the fact that *ne* is a clitic to the verb and the fact that the verb is the governor of the clause to which *ne* expresses scope indicate that the presence and function of *ne* is depending on this verb. Because negation modifies the character of its governor we decide that the relation of *ne* will be labeled as a modifying one. The dependency relations for *ne* in the examples above should therefore be the following:

For (20): **ne=MOD(pouvons)**
For (21): **n'=MOD(ont)**

Now that the scope marker of *pas* (*ne*) indicated in which clause the negation is interpreted, it is the question which part within that clause is appropriate to be the governor of the negating word *pas*. This can be revealed by the position of *pas*. *Pas* occurs at a position that can be used by adverbs to modify the inflected verb it is following. Replacing *pas* with another adverb, like *certainement* in the examples below, demonstrates this:

(22) Nous pouvons **certainement** nous permettre de perdre Paris
      We   can   certainly   us   to-allow   of to-lose Paris
      We can certainly allow ourselves to lose Paris

      ne=MOD(pouvons)

(23) Le suspens se   poursuit donc pour les 2 Français qui ont
      The suspense itself continues so   for the 2 French   who have
      certainement terminé leurs rencontres
      certainly   ended   their meetings
      So the suspense continues for the 2 French who have certainly completed their meetings

      certainement=MOD(ont)

The position of *pas* and the fact that it can also be considered an adverb (Godard, 2004), or at least has characteristics in common with adverbs (Pollock, 1989), leads to the assumption that *pas* functions as a modifier of the inflected verb it follows as well. The correct representation in dependency grammar to express this function should therefore be the same as other adverbs at this position: a modifying dependency to the inflected verb it follows:

For (20): **pas=MOD(pouvons)**
For (21): **pas=MOD(ont)**

When adverbs occur at the same time as *pas* and should be annotated as modifying dependents of the verb as well, it will not cause any conflict with the proposed dependency relations of negation because every item can have an unlimited number of dependents. (More about this in section 2.1.1.1.).

The function of the inflected verb supports treating the verb as the governor of *pas*. This is because in sentential negation the truth-value of the entire clause is negated (Jones, 1996). Because in dependency grammar an item can only have one direct governor, the only way to express an effect towards an entire clause, is to be a dependent of its head, which in the case of sentential negation is the inflected verb:

(20) Nous [VP **ne** pouvons **pas** [VP nous permettre de perdre Paris]]
(21) Le suspens se poursuit donc pour les 2 Français qui [VP **n'**ont **pas** terminé [NP leurs rencontres]]

It can be concluded that both *ne* and *pas* should be assigned dependency relations to the inflected verb of the sentence, labelled as 'modifiers'. This means that the dependency relations for the examples mentioned above will look as follows:

For (20): **ne=MOD(pouvons) pas=MOD(pouvons)**
For (21): **n'=MOD(ont) pas=MOD(ont)**

As shown above, in principle the items of negation (*ne* and *pas*) will be dependents of the finite verb. However, in some situations it is possible that a past participle is not accompanied by an auxiliary. For example the following utterance would be acceptable (ceated by replacing *jamais* from an example which comes from the corpus, and checked by a native speaker):

(24)   Le   taux de réussite du bac   général est même à   un
       The rate of success of baccalaureate general is   even   to a
       niveau **pas** atteint   876%
       level   not reached 876%
       The success rate of the general baccalaureate is even at a level that
       not reached 876%

In that case the negation will be the dependent of the item that is then the head of the phrase, the part participle:

For (24): **pas=MOD(atteint)**

2.1.1.1. Variations in the linear structure

In sentential negation, it is possible for pronouns or other modifiers (adverbs) to occur between the two negating elements *ne* and *pas*. This has however no influence on the dependency relations of *ne* and pas. As set out in the background section, dependency grammar tries to represent the syntactic structure of a sentence outside its linear representation. A change in linear representation does not need to cause a change in syntactic representation. Because pronouns and modifiers do not change the syntactic function of negation, their dependency relations do not interfere with the relations of negation. Negation should therefore still be attached to the verb while other modifiers have their own independent relations towards their governor:

(25)   Je **ne** lui   demande **pas** de retirer   sa candidature
       I *ne* him ask   not for to-withdraw his candidacy
       I do not ask him to withdraw his candidacy

       ne=MOD(demande) lui=OBJ(demande) **pas=MOD(demande)**

(26)   Ah tout bon   sens   **n'** est donc **pas** perdu à l'   appui   de cette
       Ah all   good sense *n'* is so   not lost   to the support of this
       furia anti-française
       fury anti-French

Ah not all common sense is thus lost in support of this anti-French fury

n'=MOD(est) donc=MOD(est) **pas=MOD(est)**

(27) On a   même **pas** vu   la   benne tomber
We have even   not seen the bucket to-fall
We have not even seen the bucket fall

même=MOD(pas) **pas=MOD(a)**

We mention the fact that pronouns and extra modifiers between negating elements and the verb do not interfere with the dependency relations of negation on the occasion of sentential negation but the same principle is of course valid for other forms of verbal negation (clausal negation). In order to avoid repetitions for all forms of verbal negation and all negative items we mention it only once. All possible manifestations of verbal negation must thus be considered with a variable number of pronouns and adverbs occurring between the verb and the negative items without a change in the proposed dependencies.

Another issue to address once but that applies to every form of verbal negation is that where *ne* is absent but implied, the second part of negation will express the same kind of dependency relation as it would express in the presence of *ne*. In (27) for instance, it can be seen that verbal negation is expressed only by *pas*. In spoken language this often occurs. *Ne* can however be added in front of the finite verb without a change in meaning or grammaticality. *Pas* implies the position of the implicit *ne*:

(28) On **n'** a   même **pas** vu   la   benne tomber
We *n'* have even   not seen the bucket to-fall
We have not even seen the bucket fall

The dependencies for verbal negation therefore remain the same in the presence or absence of *ne*.

2.1.2   Clausal negation

In the construction we defined (in the background section) as clausal negation caused by *pas*, (an implied) *ne* and a required *pas* occur together before an infinitive clause:

(29) Et   surtout   [$_{vp}$**ne pas** avoir   de conclusion hâtive]]
And especially *ne*   not to-have of conclusion hasty
And especially not having a hasty conclusion

The function of *ne* (whether overt or implied) when it occurs directly in front of *pas* within an infinitive clause is similar to the function it has in case of sentential negation: It is a clitic to the verb it precedes and a scope

marker for negation pointing to the clause of which it is part (Godard, 2004). Because we aim for a consistent analysis equal functionality requires equal dependency relations. Therefore, the function of *ne* is again expressed by a dependency relation to the verb it is the clitic of and that serves as governor of the clause *ne* points to:

For (29): **ne=MOD(avoir)**

To find out the syntactic role of *pas* we look at the role other elements would fulfill at the same position. Elements at the position of *pas* that would lead to a grammatical correct sentence are adverbs:

(30) Et surtout [$_{vp}$**rapidement** avoir de conclusion hâtive]
      And especially **quickly**   to-have of conclusion hasty
      And especially quickly having a hasty conclusion

  rapidement=MOD(avoir)

The adverb (here: *rapidement*) has a dependency relation as a modifier of the infinitive[1]. The fact that *pas* has characteristics in common with adverbs and occurs at the same position as adverbs modifying the infinitive gives an indication that *pas* has the same dependency relation as adverbs at it position:

For (29): **pas=MOD(appliquer)**

This treatment is indeed justified considering the fact that the infinitive is the governor of the clause that *ne pas* negates. This is relevant considering the fact that in clausal negation *pas* only functions to negate the truth value of the entire clause (Jones, 1996). The only way to express this effect in dependency grammar is to modify the governor of the clause it negates. In case of clausal negation this governor is the infinitive. The dependencies for *ne* and *pas* in clausal negation are thus both treated as modifying relations to the infinitive verb. For the example above this results in:

For (29): surtout=MOD(avoir) **ne=MOD(avoir) pas=MOD(avoir)**
       avoir=CC(et)

For the record, in dependency grammar (which is not based on linear representation) it does not matter for the annotation of negation at what position in the sentence the infinitive clause occurs. The only thing that possibly varies is the role and thereby the dependency relation of the head of this clause towards the rest of the sentence. The role of negation is still to negate the clause it is part of. The dependency relations of *ne pas* in clausal negation thus always remain the same for any function of the infinitive clause. In

---

[1]This is only the case for structures where the adverb precedes an infinitive but does not follows a finite verb. In the latter case the adverb replaces the place of *pas* where it is uncertain whether it causes clausal or sentential negation. See background section.

the sentence below for example, the infinitive clause has a different relation to the main verb (POBJ) than in the example given before (CC) but the negation relations are the same:

(31) Mais ce   qui   est euh très dangereux c'   est de [ᵥₚne pas
     But that what is   euh very dangerous that is of *ne*   not
     avoir   son attention euh sur la   route]
     to-have his attention euh on the road
     But what is very dangerous is not having his attention on the road

     de=POBJ(est) **ne=MOD(avoir) pas=MOD(avoir)** avoir=COMP(de)

Besides the construction for clausal negation shown above, an infinitive auxiliary can also occur between *ne* and *pas*, just as in sentential negation:

(32) Marie semble [ᵥₚn' avoir   **pas** aimé ce   film] (Jones, 1996)
     Marie seems   *n'*   to-have not liked this film
     Marie seems not to have like this film

This construction is reserved for a literary style and can only be performed by some verbs (Jones, 1996). Like expected, the construction has not been encountered in the corpus. However, it is not impossible that it does occur in some spoken language outside the corpus (even if this only consists of reading out loud a written text). Dependency relations for this construction must thus be proposed. Because the construction has the same function of the one discussed before; performing clausal negation, the negating items are dependents of the head of the negated verb phrase, here the auxiliary infinitive:

For (32): **n'=MOD(avoir) pas=MOD(avoir)**

2.1.3   Constituent negation

As we have just seen, (a possible) *ne* and *pas* modify a verb and in this way negate the truth-value of an entire verb phrase. However, the use of *pas* is not restricted to sentential contexts. Constituents can also be negated (Rowlett, 1998). In this case the negation does not modify the entire sentence but only a certain constituent in the sentence. In the following paragraphs all different types of constituent negation shall be discussed.

2.1.3.1. *Pas* before a noun phrase

The first type of constituent negation we discuss arises when *pas* does not express verbal negation and occurs in front of a NP constituent. Some examples are the following:

(33) [ₙₚ**Pas** question de sortir   sans   ses papier]
     No   question of to-go-out without his paper
     No question of going out without his paper

(34) [NP**Pas** moi]
      Not   me
      Not me

When *pas* expresses constituent negation and precedes a noun phrase, this entire noun phrase this negated by *pas*. Because in dependency grammar each item can only have one governor, *pas* needs to express its influence towards the entire noun phrase by a dependency relation to only one element, the head of this phrase. Because the head of the noun phrase (a (pro)noun) is the only element that is connected with all the elements within its phrase, the head is the only suitable governor for pas to represent its effect on the entire clause. The dependencies for the examples above become thus:

For (33): **pas=MOD(question)** de=MOD(question)
For (34): **pas=MOD(moi)**

Another example of constituent negation is:

(35) Mais **pas** la   sécu
      But   not the safety
      But not the safety

According to the proposed annotation we make the head of the negated noun phrase, the noun *chose*, the governor of *pas*:

For (35): **pas=MOD(sécu)**

In this example it can be seen that it is not always the case that the head of a noun phrase is the only or first element of the phrase. In a noun phrase can also occur adverbs, adjectives, determiners, quantifiers or other nouns. Because they are not the head of the noun phrase they do not play a part in the dependency relation of negation.

Pas before *de* + noun phrase

A special element that might occur in a noun phrase is *de*. It is special because it has a number of prepositional uses but in many contexts it calls for a different treatment (Abeillé et al, 2004[2]). Abeillé et al (2004) classify the use of *de* in two categories: the so-called oblique *de*-phrases where the properties of *de* can be accounted for by analyzing it as a preposition:

(36) Je me   souviens [PPde ce   film]
      I myself remember of   that film
      I remember that film

and non-oblique occrences of *de* with nouns that behave like noun phrases:

(37) Un courrier contenant [NPde la   poudre blanche suspecte]
      A letter   containing of   the powder white   supicious

17

A letter containing suspicious white powder

The non-oblique use of *de* is annotated in the corpus by a dependency relation towards the head of the noun phrase. Because *de* is part of the noun phrase there is no problem in applying the proposed annotation whenever *pas* expresses non-verbal negation and precedes a noun phrase. It is a dependent of the head of the noun phrase:

(38) Alors vous avez entendu le   garde des Sceaux [NP **pas** d'amnistie
So   you have heard   the guard of Seals   not of   amnesty
des délits   politico-financiers] d' une manière directe ou indirecte
of   ofences political-financial of a   manner direct   or indirect
So you heard the Minister of Justice no amnesty for political and financial crimes in a direct or indirect manner

d'=DET(amnistie) **pas=MOD(amnistie)**

In the case of oblique use of *de*, *de* is not part of a noun phrase but creates a prepositional phrase. Now pas no longer precedes a noun phrase but a prepositional phrase which prevents the negation of a noun phrase. In that case the annotations for negated NPs can of course not apply.

(39) Euh c' est que euh beaucoup de gens   considèrent que ça
Uh it is that uh many   of people consider   that it
dépend de l'   autre et   [PP**pas** de soi]
depends of the other and not   of self
Uh it is that uh many people consider that it depends on the other and not on themself

In section 2.1.3 the annotations for negated prepositional phrases will be discussed.

   Although the two uses of *de* cannot be distinguished by form, it should be possible to distinguish them automatically based on different POS tags ('PRP' versus 'PRP:det' for example). However, the tags in the corpus are assigned automatically and have lead to the same tag for all instances of *de*: they are tagged as preposition ('PRP'). This might cause errors for the automatic implementation of the annotations. More about the automatic annotations can be found in section 3.

Pas before a quantifier + *de*

A situation in which it is clear in what manner *de* is used without knowing its POS tag, is when it follows a quantifier, forming one of the following constructions: *pas beaucoup de* (*not a lot of*), *pas trop de* (*not too much of*), *pas tellement de* (*not so much of*), *pas peu de* (*not little of*), *pas mal de* (*some of*), *pas moins de* (*not less of*), *pas plus de* (*not more of*). See for example:

(40) Oui [_{NP} **pas** mal d'   averses encore aujourd'hui]
      Yes not bad of   heavy-rains still   today
      Yes still quite a lot of heavy rain today

      mal=DET(averses) de=MOD(beaucoup)

Like other adverbs in a noun phrase, the quantifying adverbs should be annotated as the dependent of the head (according to the annotation guide of the corpus). The non-oblique *de* is in this situation is (also according to the annotation guide of the corpus) annotated as a dependent of *mal*. Both the presence of the degree adverb as the presence of *de* in the example above do not effect the relations for negation because one item can have multiple dependents. Therefore, *pas* can still be annotated with a dependency relation to the head of the noun phrase that follows (*averses*). Although it is possible to argue for a semantic interpretation where *pas* does not negate the entire noun phrase but only the quantifier, consistent analysis regarding the syntactic annotations requires that we consider *pas* as a dependent of the entire noun phrase. This is because the quantifier is annotated as a dependent of the head of the noun phrase by means of a determiner labelled relation. A determiner is never negated directly but only occurs inside a negated noun phrase. *Pas* is therefore the dependent of the noun instead of the determiner. An example can be seen in sentence (35), which is repeated here:

(41) Mais **pas** la   sécu
      But   not the safety
      But not the safety

Because the presence of a determiner does not prevent the dependency of *pas* to the noun phrase, and quantifiers like *mal de* are labelled as determiners as well, the quantifiers should not influence the dependency of *pas* to the noun phrase either. This means that *pas* always needs to be annotated as a dependent of the head of the noun phrase that follows, regardless the presence of quantifiers. The dependency of the example give above is therefore the following:

For (40): **pas=MOD(averses)**

2.1.3.2. *Pas* before a prepositional phrase

When *pas* does not express sentential negation and directly precedes a prepositional phrase, it negates the entire prepositional phrase. It causes constituent negation of the prepositional phrase. It should therefore be annotated as the dependent of the head of the prepositional phrase, the preposition:

(42) Hein c' est sûr   que euh quand euh vous avez le   téléphone
     Eh   it is certain that euh when euh you   have the telephone
     portable vous vous avez euh la   main euh qui   est euh sur le
     portable you you   have euh the hand euh that is   euh on the
     téléphone et   [PP**pas** sur le   volant]
     telephone and not   on the wheel
     Eh it is certain that euh when euh you have your mobile phone you
     you have euh your hand euh on the telephone and not on the wheel

     pas=MOD(sur)

We repeat the fact that every type of constituent negation only takes place
when verbal negation cannot take place. An example where it is not certain
that verbal negation does not take place is the following:

(43) C'   était euh juste une une chose c' est bon   premièrement c' **est**
     That was euh just a   a   thing it is good firstly   it is
     pas par rapport à l'   internet sinon   euh par rapport une euh
     not by   report   to the internet otherwise euh by   report   a   euh
     une agence de voyage pour un un vol   euh Paris Lima
     a   agence of travel for   a a flight euh Paris Lima
     That was euh just a a thing ok first it is not about internet but about
     a travel agency for a a flight euh Paris Lima

     pas=MOD(est)

Pas occurs here before a prepositional phrase but also after an inflected
verb, *pas*. In these situations *pas* does not express constituent negation but
negates the entire sentence. The annotations for sentential negation will thus
apply.

2.1.3.3. *Pas* before an adjectival phrase

Another type of constituent that can be negated by *pas* is an adjectival
phrase. It occurs when *pas* does not express sentential negation and precedes
an adjectival phrase. *Pas* function towards the adjectival phrase needs to be
expressed again by a dependency relation to the head of the clause, here the
adjective:

(44) Eve vous aussi vous êtes plutôt contente de ces   de ces   voyages
     Eve you too   you are rather happy   of that of that travels
     [adjP**pas** trop chers]
     not   too expensive
     Eve you are also rather happy with those those not too expensive
     travels

     pas=mod(chers) trop=mod(chers) chers=MOD(voyages)

Semantically, it is clearly possible that *pas* in the above example does not actually negate the entire adjectival phrase but only the adverb *trop*. However, the negation shall be attached here with a dependency relation towards the following head, which is the head of the adjectival phrase. This is decided considering the fact that *pas* is no dependent of the quantificational expressions like *beaucoup de* in case of constituent negation of a noun phrase either. Instead, the negation is a dependent of the governor of the quantifier. Because adverbs like *trop* have semantically a similar role as quantifiers, and because they are annotated in the same way, there is no reason to treat items as *trop* differently than expressions like *beaucoup de*. The governor of the negative item in the example above is thus the head of the adjectival phrase *chers*.

2.1.3.4. *Pas* before an adverbial phrase

Pas negates an adverbial phrase in case it does not express sentential negation and is followed by an adverbial phrase that is not embedded in one of the phrases discussed in the paragraphs above. In this case, *pas* needs to express its negation of the entire adverbial phrase by its connection towards the head of the phrase, the adverb. An example is the following:

(45) Soutien au maire jusqu'à la   fin de la   mandature [_{AdvP}**pas**
       Support at mayer until   the end of the mandate   not
       plus] disent- ils
       more say   they
       Support to the mayer untill the end of the mandate not more they say

pas=MOD(plus) plus=MOD(disent)

2.1.4 *Pas* with ellipsis

For all types of negation *pas* can express, annotations were proposed with *pas* as the dependent of the item or clause it negates. However, it is not always the case that *pas* can express its sentential, clausal or constituent negating function by one of the dependency relations discussed above. Sometimes the dependency relation of *pas* is already claimed by another governor. This happens in some cases of ellipsis. Look for example at the following sentence:

(46) Dominique Vaucy chez   les voyagistes   on utilise les vols
       Dominique Vaucy among the tour-operators we use   the flights
       lowcost ou **pas**
       lowcost or not
       Dominique Vaucy at the tour operators do we take lowcost flights or not

pas=CC(ou)

Here *pas* will be annotated as the dependent of the conjunction *ou*. This is decided on the basis of two facts. First, *pas* does not occur next to a verb or left of a constituent, which makes it impossible to be attached in one of the ways we discussed above. And secondly, *ou* requires a dependent at each side to express its function as conjunction. Because *pas* is the only element to its right it is the only element that could function as the right dependent of *ou*. The annotation guide prescribes that the dependents of conjunctions are attached with a CC-label, which makes that we assign the dependency relation of *pas* the same label.

When ellipsis causes *pas* to precede a phrase but also follow the conjunction *ou*, we have to reason well to detect what is the correct governor of *pas*; *ou* or the head of the constituent it follows:

(47) On a   à la fois un soucis euh d' essayer d' essayer de
      We have of the time a   worries euh of to-try of to-try of
      comprendre   comment les choses se   sont passeés et
      to-understand how   the things themselves are   past   and
      aussi une inquiétude de voir   comment euh les personnes mises
      also a   concern   of to-see how   euh the people   put
      en examen   ou pour lesquelles une responsabilité euh nous
      in examination or for   which   a   responsibility euh we
      semble attaché   vont vont réagir et   reconnaître ou [NP**pas** leur
      seems attached go   go   react and recognize   or not   their
      responsabilités]
      responsibilities
      We have both worries about euh trying to trying to understand how
      things have happened and also concerns about seeing how the arrested
      people or those who have a responsibility euh that seem to us related
      will will react and recognize or not their responsibilities

One of the behaviours of *pas* that is presented before is that when it precedes a noun phrase it can modify this phrase. If we would choose to represent this by a modifier relation to *responsabilités* in the above example, we would have to choose another element as the right dependent of the conjunction *ou*. The conjunction would then have the dependents *reconnaître* and *leur responsabilité*. However, this does not seem correct since conjunctions normally only coordinate elements from the same POS category. Therefore we know that we are dealing with ellipsis. Knowing besides that *ou* coordinates *reconnaître*, it must also coordinate a -due to ellipsis *implicit*- verb at its right side, which is modified by *pas* ((*ne*) *pas reconnaître*). To express the implicit negation of *pas* towards the verb *reconnaître* and to express its role as a coordinated element, *pas* needs to attach to *ou* again, leading to the following dependency relations:

For (47): réagir=CC(et) et=OBJ(vont) reconnaître=CC(ou) ou=CC(et)

pas=CC(ou)

In situations where exactly the same word category occurs at the left and right side of the coordination there is no difficulty in figuring out the correct dependency relations. (Again, extra optional modifiers between these elements do not influence the proposed dependencies.) The conjunction coordinates both elements and *pas* can function as a modifier to the element it negates. In sentence (48) for example, there are adjectives before and after the conjunction while *pas* modifies one of them.

(48) Et   encore une fois il est toujours plus facile en s'
      And yet   a   time it is always   more easy in themselves
      adressant à quelqu'un quelqu'   il soit **dur** ou **pas dur** d'
      sending   to someone   whatever it or   hard or not hard elsewhere
      ailleurs euh qu' il soit dûment dûment euh   licencié pour
      euh   that it or duly duly   euh   licenced for   to-can
      pouvoir avoir un   recours éventuel
      to-have a   claim eventual
      And again it is always easier to address someone whether he is tough or not tough besides euh whether he is duly duly fired in order to do an eventual claim

      'first' dur=CC(ou) ou=OBJ(soit) **pas=MOD('second' dur)** 'second' dur=CC(ou)

The consideration as to whether *pas* is a dependent of a conjunction or another element only plays a part when ellipsis is performed with the conjunction *ou*. Because although the construction of an implicit category after *pas* might be grammatically possible for all types of coordinations, semantically it will only occur with *ou*:

- On utilise les vols lowcost ou pas
  utilise=CC(ou) pas=CC(ou) on=SUJ(ou)

- *On utilise les vols lowcost et pas

- *On utilise les vols lowcost mais pas

It can be seen that *pas* cannot follow *et* or *mais* on its own. Therefore, in grammatical sentences constructed with conjunctions other than *ou*, the governor of *pas* is not *ou*. The negation can modify some other element in the way described in the previous paragraphs:

(49) On utilise les vols   lowcost et   (ne) (utilise) pas les autres
      We use   the flights lowcost and *(ne)* (use)   not the other
      vols
      flights
      We take the lowcost flights and do not take the other flights

utilise=CC(et) prend=CC(et) **pas=MOD('second' vols)**

(50) On utilise les vols lowcost mais (ne) (utilise) pas les autres
We use the flights lowcost but *(ne)* (use) not the other
vols
flights
We take the lowcost flights but do not take the other flights

'first' utilise=CC(mais) 'second' utilise=CC(mais)
pas=MOD('second' vols)

It should be mentioned that there is a notable difference between the annotations of *pas* occurring after *ou* and the annotations of *pas* described before. This difference lies in the way to automatically implement the annotations. Whether *pas* is the dependent of *ou* or of the clause that follows is namely the only thing necessary to implement annotations that cannot be estimated based on information about word forms and POS tags only. Here, also information about the existing dependency relations is necessary so the POS of the left dependent of *ou* can be retrieved. On the basis of that and the POS of the words to the right of *pas*, it can be decided of what element *pas* is the dependent. More details about the automatic implementation of the proposed annotations and the exact rule to attach *pas* occurring after *ou* can be found in section 3.

2.1.5 *Pas* in a separate sentence

In the previous chapters we have seen all kinds of situations in which (*ne*) *pas* occurs, causing sentential, clausal or constituent negation. In the remaining cases, *pas* does not negate anything in its own sentence. An example can be seen below where *pourquoi pas* is like an additional remark after the closure of a former sentence, although this might be difficult to see in a corpus without sentences boundaries. *Pas* refers here to the entire previous sentence and has therefore no governor in its own sentence. Because syntactically we only consider dependencies within sentences, we will leave the mentioned occurrence of *pas* unattached:

(51) Un sixième maillot blanc à poids rouges pour Virenque histoire
A sixth shirt white to weight red for Virenque story
de rejoindre pourquoi **pas**
of to-join why not
A sixth white shirt with red dots for Virenque in order to join, why not

Pourquoi=MOD(pas)

The reason why we make *pas* the governor of *pourquoi* instead of the other way around is because we have no linguistic reason to prefer one direction

over the other so we can choose for a maximum consistency of the dependency relations. We do not want to invent new types of annotations when this is not necessary. Besides, *pourquoi* being the dependent of *pas* is exactly what is prescribed in the existing annotation guide of the corpus. This provides *pas* the opportunity to be the dependent of another item in case it does negate a phrase in the same sentence. This can be seen for example in (52) where *pas* can be annotated according to the prescribed annotations for clausal negation:

(52) Pourquoi **ne pas** acheter un billet euh hormis le   prix
     Why   *ne* not to-buy a   ticket euh except the price
     évidemment
     obviously
     Why not buying a ticket euh except for the price obviously

pourquoi=MOD(pas) **ne=MOD(acheter) pas=MOD(acheter)**

It can be concluded that when *pas* occurs in a different sentence than the item(s) to which it refers, (as in (51)) *pas* has no governor.

2.1.6   *Non Pas*

Now all situations in which *pas* can occur and its corresponding dependency relations are discussed. In order to keep the analysis of *pas* consistent with the proposed annotations, fixed expressions that occur with pas and change the nature of *pas*, must be annotated in such a way that they do not conflict with the proposed dependencies of pas.

   The first item that has been seen often in combination with *pas* is *non*, as in:

(53) Et   il est bon que le   premier ministre fasse que ce   droit
     And it is   good that the first   minister does that this right
     soit   opérationnel et **non pas** clandestin comme c 'est le
     whether operational   and not not clandestine as   it is   the
     cas   aujourd'hui
     case today
     And it is good that the prime minister makes this right operational and not clandestine as is the case today

Because *non* is also a negative item on its own meaning *not* we must see whether in combination with *pas* it is a fixed expression belonging to *pas* or if it is some item that is not related to *pas* but has its own function in the sentence. It can be seen that in all cases where *non* precedes *pas*, *non* can be omitted without losing any structure or meaning in the sentence. Besides, it always occurs at the direct left side of *non*. Together they are used exclusively to negate a second part of a comparison (X and not Y). Therefore, *non* can be attached as a multiword dependent of *pas*. Multiword

dependents are namely designed for a fixed sequence of words expressing one fixed meaning together. Making *non* the multiword expression of *pas* instead of the other way around, ensures consistency: we do not need extra rules to handle *pas* and others to detect the governor of *non*. Instead, *pas* keeps functioning in the same way as when it would be alone. Making *non* the multiword dependent of *pas* leads to the following dependencies for the above sentence:

For (53): **non=MULT(pas) pas=MOD(clandistin)**

2.1.7 *Pas Du Tout, Pas Non Plus*

Just like *non pas* there are more combinations with *pas* that need multiword expressions. In this case not because their meaning and function is equal to *pas* alone, but just to preserve consistent dependencies. This is the case for *pas du tout* (*not at all*), and *pas non plus* (*neither*):

(54)   Et   je vous assure que c' est **pas du tout** dans notre volonté
       And I you assure that it is not of all   in   our   will
       And I assure you that it is not at all our will

(55)   Par example oui ça   ça   aurait été une une posibilité si on ne
       By example yes that that would was a   a   possibility if we (*ne*)
       veut **pas non plus** un changement peut-être trop brutal
       can not not more a   change   maybe   too brutal
       *For example yes that would be a possibility if you don't want either a
       perhaps too brutal change*

According to the proposed dependencies for constituent negation, *pas* would be assigned dependency relations to the preposition *du*, and adverb *plus* (pas=MOD(du) and pas=MOD(plus)). However, there are no linguistic clues in the structure or meaning of the word sequences that support such annotations. This makes it also plausible that the last parts are dependents of the negation instead of the other way around (du=MULT(pas) and non=MULT(pas)), creating a fixed expression with *pas*. In that case *pas* will be the governor of the expressions and have room to modify parts outside the fixed expression in the ways described in the previous paragraphs. This leads to:

For (54): **pas=MOD(est)**
For (55): **pas=MOD(veut)**

The entire expression now serves as a special type of negation (*not at all* or *neither*) modifying some other part of the sentence in the same way as *pas* would alone. If another part of the sequences would be chosen as the governor, we would have to treat this governor in such a way that it can express the function of negation as well. More easy and consistent is it to

keep *pas* the head of the entire negating expression leading to the following multiword dependencies for instances of *pas du tout* and *pas non plus*:

For (54): **du=MULT(pas) tout=MULT(du)**
For (55): **non=MULT(pas) plus=MULT(du)**

2.2 (Ne) Plus

Another negative item for which dependency relations shall be proposed is *plus*. In verbal negation it is used in combination with *ne*, forming a meaning like *no more* or *not anymore*. The contribution of *ne* is the same as discussed for other items expressing negation: it is the scope marker of the negation but can be left out in colloquial use.

An important characteristic of *plus*, in contrast to *pas* for example, is that it needs an explicit or implied first pair part in order to have a negative connotation. Without this (implied) element *plus* has a positive meaning: *more*. This means that *plus* alone cannot express constituent negation. Because in this project only annotations for negations are proposed, *plus* shall only be dealt with when it has a negative meaning, in which case it expresses verbal negation.

Sometimes it is ambiguous whether *plus* expresses (verbal) negation or not. Because a positive meaning of *plus* might need another treatment than the negative one (see section 2.2.1) there must be a way to decide in which case it shall be treated positively and in which case negatively.

The ambiguity about the meaning of *plus* arises whenever *plus* occurs without the scope marker *ne* in a construction that is used for verbal negation. This is because in such a construction *ne* can be implied or not leading to an ambiguity as to whether *plus* expresses verbal negation or has a positive interpretation. The sentence below for example, can be interpreted in two ways.

(56) Je veux plus de pain
I   want more of bread
I want more bread or *I want no more bread*

This ambiguity of whether or not *plus* has a negative meaning in sentences like (56) arrises only in the transcription of the spoken utterances. In speech both uses can be distinguished by their pronunciation; in the positive use of *plus* the last *s* is pronounced while in the negative use the *s* is silent. However, this information is lost in the transcription where both pronunciations, and thus both uses of *plus* are written in the same way. Because we have only access to transcriptions we cannot distinguish automatically positive uses from negative uses of *plus*. It must thus be decided beforehand what to do with the ambiguous cases: treating them as negation or not. It is decided that *plus* without scope marker will be not be treated as negative item because in the corpus there are only two cases found where negation is performed by *plus* without *ne* (according to native speaker judgements) of which one could be the result of a wrong transcription. The possible wrong transcription occurs in the following example:

(57) Et   puis quand on a   lévé les yeux on a   vu que la
And then when we have lifted the eyes we have seen that the

      benne était **plus** sur le   cable
      bucket was more on the cable
      And then when we looked up we have seen that the bucket was not on
      the cable anymore

In this utterance it cannot be heard if there occurs a *ne* before *était* or not. The pronunciation is the same in both cases which might have caused a wrong transcription.

Except for this example and one other transcription, in all cases where *plus* occurs without *ne*, *plus* has a positive meaning, as interpreted by a native speaker. The frequency of the negative uses of *plus* without *ne* lead to the decision that they will not be taken into account in the automatic implementation of the proposed annotations. It will be assumed that *plus* only expresses verbal negation when it co-occurs with *ne*. This consideration will lead to some errors in the automatic implementation of the annotations but due to the low frequency of verbal negation without *ne*, the amount of errors shall be minor. For manual annotation dependency relations for *plus* in isolation will be proposed such that they can be applied according to the interpretations of native speakers, but these annotations will not be taken into account in the automatic implementation.

Regarding constituent negation, it is mentioned above that *plus* alone is positive which makes it impossible to express constituent negation on its own. However, sometimes the item is accompanied by the negative *non*. In that case, *non plus* does perform negation to a certain constituent (see paragraph 2.2.3). Therefore, annotations shall be proposed for *plus* in combination with *non*.

Below annotations shall be proposed for all the occurrences of *plus* that are just mentioned.

2.2.1   Sentential negation

Sentential negation by *plus* is characterized by *ne* before and *plus* after an inflected verb. An example from the corpus is:

(58)   Mais elle [_{VP}**ne** travaillera **plus** toute seule dans son canton]
        But she *ne*   will-work more whole alone in   her canton
        But she will not work anymore completely alone at her canton

Because *ne* has the same role for every type of negation in which it occurs, being a clitic to the verb and a scope marker for negation, it is assigned the same dependency relation as when it occurs in combination with *pas*:

For (58): **ne=MOD(travaillera)**

To see what item is the correct governor of *plus* we first look at its position. Sentence (59) shows that the position of *plus* is used by adverbs to modify the preceding verb:

(59)   Mais elle [$_{VP}$travaillera **certainement** toute seule dans son
But   she will-work   certainly   whole alone in   her
canton]
canton
But she will work certainly completely alone at her canton

certainement=MOD(travaillera)

Because *plus* has characteristics in common with adverbs, it is likely that *plus* modifies the verb here as well. In that case it should be assigned the same dependency relation as other adverbs (see (59)). When both other adverbs and *plus* occur after the verb they can both be dependents of the verb because any item can have an unlimited number of dependents. *Plus* as a dependent of the verb is indeed the appropriate representation considering the fact that *ne plus* in sentential negation applies to the entire sentence. This effect can only be represented by a dependency relation to the head of this entire sentence which is the inflected verb. *Plus* in sentential negation is thus assigned the same dependency relation as *pas* in sentential negation:

For (58): **plus=MOD(travaillera)**

When *plus* occurs without its scope marker, expressing sentential negation as well (judged by native speakers) it will be annotated in the same way:

(60)   Et   puis quand on a   lévé les yeux on a   vu   que la
And then when   we have lifted the eyes   we have seen that the
benne était **plus** sur le   cable
bucket was   more on   the cable
And then when we looked up we have seen that the bucket was not on the cable anymore

plus=MOD(était)

However, as mentioned before, this annotation can only be added by human annotators who in such situations, can make the difference between negative and positive uses of *plus*. This annotation shall not be taken into account in the automatically assigned annotations. Instead for the automatic implementation of the annotations *plus* will be considered positive here, and requires therefore annotations corresponding to the positive function of *plus*. Although these annotations are not part of the annotation scheme described here, we mention that the positive *plus* can function as quantifier and should therefore be a modifying dependent of the head at its right side. For the above example this would be: plus=MOD(sur). This annotation is not the same as the negative annotation of *plus* which is why a decision has been taken as to how ambiguous cases should be treated.

2.2.2 Clausal negation

Clausal negation caused by *ne plus* is presented in sentences like:

(61) ...et   fort   son intention de [_{VP}**ne plus** venir   sur le   Tour de
      ...and strong his intention of *ne*   more to-come on the Tour de
      France]
      France
      ...and strongly his intention of not coming anymore to the Tour de
      France

Here, *ne plus* negates the verb phrase that follows. This is indicated by the scope marker of the negation. To represent this effect the negative items will be attached as dependents of the infinitive. Besides, this representation is consistent with clausal negation by *ne pas*:

For (58): **ne=MOD(venir) pas=MOD(venir)**

Another possible construction for clausal negation is mainly used in literary style and is also used for clausal negation by *pas*. It has not been encountered in the corpus but annotations need to be proposed for this construction too because it is not completely impossible that they would occur in a very large corpus of spoken text. The following example is a variation to 32 which comes from (Jones, 1996) and is checked by a native speaker:

(62) Marie semble [_{VP}**n'** avoir   **plus** aimé ce   film]
      Marie seems   *n'*   to-have more liked this film
      Marie seems not have like this film

      n'=MOD(avoir) plus=MOD(avoir)

In this situation the head of the entire sentence that is negated is the infinitive verb (here *avoir*). Therefore, this item will be considered the governor of the verb. This annotation corresponds to that of *pas* in the same construction and to the other manifestation of clausal negation caused by *plus*.

Also for clausal negation it is the case that when *plus* would occur without its scope marker *ne* and human annotators judge that it still expresses clausal negation, it needs to be annotated with the same dependency relation as *plus* within the presence of *ne*. Again this annotation shall not be implemented automatically.

2.2.3 *Non Plus*

A variation to the negation (*ne*) ... *plus* is (*ne*) ... *non plus*. The meaning of this construction differs in some situations from the meaning of (*ne*) *plus*; in some situations it expresses *not either* instead of *not anymore*. In the following sections all manifestations of *non plus* shall be discussed and corresponding annotations shall be proposed.

2.2.3.1. Sentential negation

An example with *non plus* from the corpus where it expresses sentential negation is the following (variation to a sentence from the corpus, checked by a native speaker):

(63) Elle ne dit **non plus** mot
      she *ne* says not more word
      She did not say a word either

The word *non* is a fixed item used to distinguish the meaning of *ne ... non plus* (here: neither) from *ne ... plus*. Besides, *non* occurs in combination with *plus* always directly to the left of this item. Because of these reasons it will be annotated as a dependent of *plus* with a multiword label. In that way, *plus* can be the element to represent the effect of the expression to other parts in the sentence. This treatment corresponds to the dependency relations of the fixed negation *ne pas* (section 2.1.6). For sentence (63) where sentential negation takes place, *plus* can thus be annotated in the way already proposed for *plus* in sentential negation while *non* is the dependent of *plus*:

For (63): **ne=MOD(dit) non=MULT(plus) plus=MOD(dit)**

Because *non plus* can only be interpreted negatively it can occur in spoken language without *ne* without causing an ambiguous interpretation. There are no examples from this in the corpus but, based on native speaker judgement, we can state that example (63) without *ne* would be acceptable in spoken French and still have a negative meaning. Because *non pas* is always negative, annotations for *non plus* without *ne* can also be automatically implemented without problems.

2.2.3.2. Clausal negation

Clausal negation by *non plus* occurs in two different constructions. First, it can be structured as *non plus* preceding an infinitive as shown in the following example (variation to a sentence from the corpus, checked by a native speaker):

(64) Mais ce   qui   est euh très dangereux c'   est de [_{VP}**non plus**
      but that what is euh very dangerous that is of not   more
      avoir   son attention euh sur la   route]
      to-have his attention euh on the road
      But that what is very dangerous is not having his attention on the road anymore

And secondly, it is possible that *non plus* follows the infinitive (checked by a native speaker):

(65) Mais ce   qui   est euh très dangereux c'   est de [$_\text{VP}$(**n'**) avoir
but   that what is   euh very dangerous that is   of (*ne*)   to-have
non plus son attention euh sur la   route]
not   more his attention euh on the road
But that what is very dangerous is not having his attention on the
road anymore

The presence of *ne* is optional in (65) and not possible in (64). Both sentences
have however scope over the same phrase. This can also be seen at the equal
translations. In both sentences the negation negates the phrase in which it
occurs, namely a verb phrase. It can therefore be handled as a dependent of
the head of the negated verb phrase, the infinitive. *Non* is in line with other
occurrences of *non plus* annotated as multiword dependent of plus. This
leads to the following dependencies:

For (64): **non=MULT(plus) plus=MOD(avoir)**
For (65): **(ne=MOD(avoir)) non=MULT(plus) plus=MOD(avoir)**

Because there are no constructions where an infinitive occurs before as well
as after *non plus*, there is no problem to automatically implement only one
of the above annotations (see section 3).

2.2.3.3. Constituent negation

As mentioned before, *plus* alone does not express constituent negation. How-
ever, the combination of *non* with *plus* is negative and able to negate con-
stituents. The constituent it negates can occur to its left as well to its right.
This can be seen for instance in the following examples that are checked by
a native speaker:

(66) Le bilan   que vous avez présenté n'   a   pas pris   en compte
the results that you have presented (*n'*) have not taken in account
les humiliations quotidiennes. [$_\text{NP}$Les vies gâchées **non plus**]
the humiliations daily   the   lives wasted not more
The result that you have presented did not take into account the daily
humiliation. The destroyed lives not either.

(67) Nous avons affiché   dans la   colonne de droites les vies perdues
we   have displayed in   the column of right   the lives lost
et   [$_\text{NP}$**non plus** les vies gâchées].
and not   more the lives wasted
we have displayed in the column to the right the lost lives and not
anymore the wasted lives.

Both negations have a different meaning. In (66) *non plus* expresses *not*
either, and in (67) it means *not anymore*. However, in both constructions

non plus negates the phrase to which they belong, here NPs. The head of the negated phrases will serve again as the governors of the negation while *non* is consistently a multi dependent of *plus*, leading to the following dependency relations:

For (66): **non=MULT(plus) plus=MOD(vies)**
For (67): **non=MULT(plus) plus=MOD(vies)**

Other types of phrases than NPs can be negated by *non pas* too. They can be treated equally as the examples above. In every form of constituent negation *plus* shall be attached as a dependent of the head of the largest phrase it negates while *non* is its multiword dependent of *plus*. Other constituents that can be negated by *non plus* are for example (constructed and checked by a native speaker):
prepositional phrases:

(68) A   t' on au moins retrouvé ses empreintes sur le   volant? [PPSur
      has *t'* we at less   found   his fingerprints on the wheel   on
      le   volant **non plus**]
      the wheel not more
      Are at least the fingerprints found on the wheel? Not on the wheel either

      non=MULT(plus) plus=MOD(sur)

(69) Les commandes sont sur le   levier   de vitesse et   [PP**non plus**
      the orders   are on the handle of speed and not   more
      sur le   volant]
      on the wheel
      The gears are on the gear shift and not on the wheel anymore

      non=MULT(plus) plus=MOD(sur)

   adjectival phrases:

(70) Ça   ne sera   pas trop cher   non plus comme cadeau pour
      That *ne* will-be not too expensive not more like   gift   for
      lui? [adjPTrop chers   **non plus**]
      him too   expensive not more
      Will that not be too expensive as a gift for him? Not too expensive either

      non=MULT(plus) plus=MOD(chers)

(71) La location d' appartement en proche banlieu   est raisonable et
      the rental   of apartment   in close   suburban is reasonable and
      [adjP**non** *plus* trop   cher comme à   Paris]
      not   more expensive like to   Paris

> *The rental of the apartment in the close suburbans is reasonable and not so expensive anymore as in Paris*

and adverbial phrases:

(72) Donc il y a   beaucoup plus de sucre que dans les M&Ms?
      so   it *y* have much   more of sugar than in   the M&Ms
      [_{advP}Beaucoup plus **non plus**]
      much   more not   more
      so it contains much more sugar than M&Ms? Much more neither

(73) Il n y a   pas tellement moins de pollution à Nancy qu' à
      it *n' y* have not so   less   of pollution to Nancy than to
      Strasbourg et   [_{advP}**non plus** beaucoup plus qu'a Epinal]
      Strasbourg and not   more more   more than to   Epinal]
      There is not so less pollution at Nancy than at Strasbourg and not anymore a lot more at Epinal

Manually it can easily be detected from the meaning of the sentence at which part of *non plus* the negated constituent occurs and therefore, to which item the dependency relation of *plus* should attach. Automatically however, there needs to be a way to decide whether the negated constituent occurs to the left or right side of *non plus*.

Non plus in its meaning of *not anymore* occurs before the negated constituent and is always used to express a contrast (to a situation before). The other use of *non plus* (not either) does not do this. On the contrary, it expresses agreement with another part in the text. Therefore, it can be assumed that when *non plus* occurs directly after a conjunction (*mais* or *et*) or at the beginning of a sentence it is probable that the constituent that *non plus* negates is the phrase following *non plus*. This might not give a 100% accuracy but should cover most of the cases correctly. More details about the automatic implementation of the proposed annotations can be seen in section 3.

All constructions of *non plus* are now discussed. Constructions that are not discussed here but which might be possible with another negation, like *pas* for example, are not acceptable with *(ne) (non) plus*. For example *non plus* as only items after *ou* is not possible, although it is for *pas*:

(74) *As   tu   travaillé plus tôt   ou **non plus**
      Have you worked   more early or not either

For all occurrences of the negative *plus* annotations have been proposed. Not only for when it expresses verbal negation but also when it expresses constituent negation with the help of *non*.

2.3 (Ne) Jamais

The last negative item for which annotations will be proposed is *jamais*, meaning *never*. *Jamais* does not need another item to be negative but has always a negative interpretation. It can express verbal as well as constituent negation and occurs typically at the same positions as *pas*. However, *jamais* has also some distributions that are not shared by *pas*. Here all situations in which *jamais* can occur will be discussed and corresponding dependency relations shall be proposed.

2.3.1 Sentential negation

Sentential negation is performed in the same way as sentential negation caused by *ne pas*:

(75) Le premier présenté en 97 **n'** a **jamais** été appliqué
     The president presented in 97 *n'* has never was applied
     The first presented in 97 has never been applied

(76) C' est **jamais** qu' une règle commune hein euh bon
     It is never that a rule common he euh good
     It is never just a common rule he euh good

The (implied) *ne* is again a clitic to the verb and scope marker of negation, while *jamais* causes negation of the entire sentence at a position that adverbs use to modify the verb. The functions of these elements are thus equal to those of *ne* and *pas* and will therefore also be annotated with the same dependency relations:

For (75): **n'=MOD(a) jamais=MOD(a)**
For (76): **jamais=MOD(est)**

A variation to the above construction for sentential negation, and a difference in comparison to the other negative words discussed, is that *jamais* can also be placed at the beginning of the sentence (Jones, 1996). An example from the corpus demonstrates this:

(77) **Jamais** un chef d' Etat étranger **n'** a été reçu avec autant
     Never a chef of state foreign *n'* has been received with as
     de chaleur d' égards et d' intimité
     of warmth of respects and of intimacy
     Never a head of a foreign state has been received with such warmth respect and intimacy

From the meaning of the sentence above it can be derived that also in such constructions the truth value of the entire sentence is negated. Besides, the scope marker indicates that the negation indeed takes place in the main clause. This construction has thus exactly the same function as when *jamais*

would occur after the finite verb. We are still dealing with sentential negation and shall therefore be consistent in the dependency relations by making the negating items dependents of the verb as well when *jamais* occurs before its scope marker:

For (77): **jamais=MOD(a) n'=MOD(a)**

Knowing that *ne* can be absent in familiar use and that *jamais* is always negative leads to the assumption that it is also possible that *jamais* occurs somewhere before the verb without the presence of *ne*. This possibility is not encountered in the corpus but is confirmed by a native speaker. In that case negation will be expressed by a single dependency relation going from *jamais* to the verb (as seen in (77)).

Sometimes a past participle is not accompanied by an auxiliary. This happens for instance in example (78). In that case the negation will be a dependent of the past participle which functions here as the head of the verb phrase that *jamais* negates:

(78) Le   taux de réussite du bac   général est même à un
      The rate of success of baccalaureate general is   even   to a
      niveau [VP**jamais** atteint 876%]
      level   never   reached 876%
      The success rate of the general baccalaureate is even at a level that never reached 876%

      jamais=MOD(atteint)

2.3.2 Clausal negation

Clausal negation by *jamais* has exactly the same syntactic structure as clausal negation by other items of negation. It will therefore receive the same annotations. An example is the following:

(79) La   sécurité routière demande à ses commerciaux à ses euh
      The security road   asks   to its merchants   to its euh
      salaries   de **ne jamais** utiliser le   telephone en   conduisant
      employers of *ne* never   to-use the telephone during driving
      The road safety asks his merchants his employees never to use the telephone while driving

      ne=MOD(utiliser) jamais=MOD(utiliser)

The literary style construction of clausal negation is also possible for *jamais*. The annotations will be the same as for the other manifestation of clausal negation:

(80) Marie semble [VP**n'** avoir   **jamais** aimé ce   film] (Jones, 1996)
      Marie seems   *n'*   to-have never   liked this film

Marie seems never have like this film

n'=MOD(avoir) jamais=MOD(avoir)

2.3.3 Constituent negation

Forms of constituent negation by *jamais* are not encountered in the corpus but they can exist. When a verb phrase is outside the scope of *jamais*, (embedded) utterances like the following can be a plausible result in spoken language (checked for acceptability by a native speaker):

Jamais before a noun phrase:

(81) L' a t on aidé durant sa traversée du désert? **Jamais** le
him have *t* we helped during his crossing of dessert never the
moindre soutien
lesser support
Did they help him during his strugles? Never any support

Jamais before a prepositional phrase:

(82) Prendra t il des vacances cette année? **Jamais** avant la fin de
will-take *t* he of holiday this year never before the end of
son mandat
his term
Will he take a holiday this year? Never before the end of his term

Jamais before an adverbial phrase:

(83) Il est déja été malade. **Jamais** très mal
he is already was sick never very bad
He has already been sick. Never very bad

Jamais before an adjectival phrase:

(84) Ça ne sera pas trop cher comme cadeau pour lui?
that *ne* will-be not too expensive as gift for him
Jamais trop chers
never too expensive
Will that not bee to expensive as a gift for him? Never too expensive

In every form of constituent negation *jamais* negates the phrase that follows. The head of this phrase will therefore be the governor of the negation:

For (81): **jamais=MOD(soutien)**
For (82): **jamais=MOD(avant)**
For (83): **jamais=MOD(mal)**
For (88): **jamais=MOD(chers)**

2.3.4 *Jamais* with ellipsis

When *jamais* is not in the correct position to modify a verb due to ellipsis, there are two possiblities for its dependency relation. Look for this at the following constructed examples (which are checked by a native speaker):

(85) C 'est maintenant ou **jamais**
      It is   now   or never
      It is now or never

(86) Les examens qu' il donne sont faciles ou jamais terminables?
      the exams   that he gives are   easy   ore never terminable
      The exams that he gives are they easy or never terminable?

In (86) *jamais* negates the constituent that follows in the way described for constituent negation. In (85) however, there is no constituent after *jamais* that it can negate. Instead, it can only function as the dependent that *ou* requires. Ellipsis can thus cause in different constructions, different dependency relations:

For (85): **jamais=CC(ou)**
For (86): **jamais=MOD(terminables)**

As in the case of ellipsis where *pas* or *non plus* are involved, the automatic implementation of the annotations need earlier annotations to retrieve the left dependent of *ou*. Based on the POS tag of this element, it can be seen whether *jamais* is a dependent of *ou* as well or if it negates a constituent to its right (see paragraph 2.1.4).

2.3.5 *Jamais* as complement

In the last function *jamais* can have, it does not negate anything but fulfills the role of complement of another item. Such a construction from the corpus is:

(87) N' avais- je pas raison de dire   que les valeurs familiales sont
      N' had   I not right   of to-say that the values family   are
      aussi solides que **jamais**
      also   solid   than never
      Wasn't I right in saying that the family values are also solid as never before

It can be seen that there is no verb or constituent phrase at a position suitable to be negated by *jamais*, just as was the case in (85). Again, *jamais* does not negate anything but functions instead only as the expression of the concept *never*. In this circumstance *jamais* can therefore not have a dependency relation to a governor it negates but its dependency relation is enforced by another element. Next to *ou*, such an element can be *que*. It occurs to the

left of *jamais* and requires a complement to its right which role can only be taken by *jamais*. This makes dependency relation of *jamais*:

For (87): **jamais=COMP(que)**

A similar example is:

(88) Plus que **jamais** Virenque sera   l'   homme à abattre au sens
      More than never   Virenque will-be the man   to to-kill   at sense
      propre comme au sens   figuré
      own   as   at sense figurative
      More than ever Virenque will be the man to beat/kill in literal as well
      as figurative sense

With this last use all the circumstances in which *jamais* can (acceptably) occur are discussed. For all of them corresponding dependency relations are proposed, which are consistent with those for the other negative items *pas* and *jamais*.

3   Rules for Automatic Implementation

In the former sections annotations have been proposed for the negative *pas*, *plus*, *jamais*, their scope marker *ne* and its multiword dependents. Here automatic implementations of these annotations will be proposed.

The implementations will be rule based. This approach has some advantages over a parser based on machine learning. For example, there is no extra corpus necessary to train the parser, which in the case of spoken French does not exist. With rule based parsing it can also be reasoned why the items are annotated in the way they are. This insight is not so straightforward in machine learning where the parser is trained on the combination of all the annotations at once. Besides, if changes in the annotation scheme are required it is easy to adapt only one rule specifically instead of needing to retrain the entire parser.

The rules for automatic annotation will be based mainly on word forms and POS tags. This is because this information is more reliable than information about previous dependency relations from the corpus. Only two rules are based on information about previous dependencies but these serve mainly as illustration about how dependency relations could contribute if they were reliable. In the tables on the next pages it will be indicated which ones use such information. Because these examples occur hardly in the corpus they shall not drastically effect the results of the automatic implementation. Other rules then these two could benefit from previous dependency relations as well but must for now be based on word forms and POS tags only.

As seen during the development of the annotation scheme, negation often has specific syntactic structures in which it occurs. Therefore, the annotations can be automatically implemented by referring to these structures. The structures concern things like *plus* after an inflected verb or *pas* before a prepositional phrase. This environment then decides what dependency relation should apply. To find the specific elements that need a dependency relation it can be reasoned where in the sequence of words and POS tags it should occur. If the governor of a negation should be the head of a phrase for example, this head can be found by exploring all possible combinations to form such a phrase and then reason for every combination what item would be its head.

On the following pages all the developed rules are presented. The second and third column represent the goal of the rule by means of an example, while the righter columns show the pattern on which the rule is based an the annotation it will add. The order of the rules in the tables demonstrates the hierarchy in which they should be implemented. When a given structure does not correspond to the conditions for the first rule for example, the conditions for the second rule can be checked, etcetera. Only in this order the correct candidates for the dependency relations can be found and multiple dependencies will be avoided. This means that the order in which the an-

notations are presented in the previous chapters does not correspond to the order of the rules for the automatic implementation of these annotations.

The rules in the following tables are demonstrated by regular expressions. In order to prevent unreadable rules one item has not been written out in the regular expressions but should be implemented. That is the fact that between all the items an undefined number of disfluencies (tagged with 'INT') is anticipated.

Sometimes rules are based on a sequence of items that is richer than that what can occur in natural language. For example, in order to attach a dependency relation to a verb an unlimited number of adverbs between the dependent and the verb is anticipated. However, it will never occur that there will be 10 adverbs between both items. Only 3 for example. More specifics about the sequence to which the rules apply is in these cases not necessary. The dependency relations connect the same items whether there occur 10 or 3 adverbs between the dependent and governor. The advantage of a sequence that is richer than what occurs in natural language is that it simplifies the rules and keeps them manageble.

Some remarks in order to read the sequences to which the rules apply: word forms are indicated by italics, POS tags are capitalized but their sub-categories are not. All the POS categories are indicated in the way they appear in the corpus, except for N, which in the corpus is NAM and NOM.

After presenting the automatic implementations on the following pages, an evaluation of the implementations shall follow in section 5.

3.1 (ne) pas (pas=ADV)

Rulename	Example of the situation	Goal	Applies to this pattern[23]	Action
Non Pas	... que ce droit soit opérationnel et non pas clandestin ...	Non=MULT(pas)	*non pas*	Non=MULT(pas)
Pas Du Tout	Et je vous assure que c'est pas du tout dans notre volonté	du=MULT(pas) tout=MULT(du)	*pas du tout*	du=MULT(pas) tout=MULT(du)
Pas Non Plus	Par example oui ça ça aurait été une une posibilité si on ne veut pas non plus un changement peut-être trop brutal	non=MULT(pas) plus=MULT(non)	*pas non plus*	non=MULT(pas) plus=MULT(non)
Ou Pas[4]	.... vont réagir et reconnaître ou pas leur responsabilités	pas=CC(ou)	POS-of-left-dependent-*ou* .* *ou* (*non*) *pas* .* ^POS-of-left-dependent-*ou*	pas=CC(ou)
(Ne) V Pas	Le suspens se poursuit donc pour les 2 Franais qui n'ont pas terminé leurs rencontres	n'=MOD(est) pas=MOD(est)	(*ne* | *n'*)? (PRO | ADVˆ(*ne* | *n'*))* Vˆ(pper | inf) (PRO | ADVˆ(*ne* | *n'*)) *pas*	(ne|n'=MOD(V)) pas=MOD(V)
(Ne) Pas Inf/Pper	Plusieurs pays à commencer par la France ont demandé à la Turquie de ne pas appliquer la sentence	ne=MOD(appliquer) pas=MOD(appliquer)	(*ne* | *n'*)? *pas* (PRO | ADVˆ(*ne* | *n'*))* V(inf | pper)	(ne|n'=MOD(V)) pas=MOD(V))

	Le taux de réussite du bac général est même à un niveau pas atteint 876%	pas=MOD(atteint)		
Pas NP	La Pologne a bien essayé de placer son ancien Premier ministre mais pas le moindre soutien	pas=MOD(soutien)	*pas* (*du tout* | *mal* (*de* | *d'*) | *beaucoup* (*de* | *d'*) | *peu* (*de* | *d'*) | DET | PRP:det | (*de* | *d'* (^PRP)) | NUM | ADJ | ADV | PRO:DEM)* (NAM | NOM) | PRO	pas=MOD((NAM | NOM) | last PRO)
	Alors ces querelles prendront-elles fin un jour pas cette semaine c'est sûr	pas=MOD(semaine)		
	Pas moi	pas=MOD(moi)		
Pas PP	... ne pouvait pas être poursuivi dans le cadre des affaires le concernant quand il etait maire de Paris en tout cas pas avant la fin de son mandat	(Second) pas=MOD(avant)	*pas* (*du tout* | ADV)* PRP	pas=MOD(PRP)
Pas AdjP	Eve vous aussi vous êtes plutôt contente de ces de ces voyages pas trop chers	Pas=MOD(chers)	*pas* (*du tout* | ADV)* ADJ	pas=MOD(ADJ)

Pas AdvP	Soutien au maire jusqu'à la fin de la mandature pas plus disent-ils	pas=MOD(plus)	*pas (du tout* | ADV)* ADV	pas=MOD(last ADV)
(Ne) Inf Pas	Marie semble n' avoir certainement pas aimé ce film	n'=MOD(avoir) pas=MOD(avoir)	*(ne* | *n')*? (PRO	ADVˆ(*ne* | *n'*))* V(inf) (PRO	ADVˆ(*ne* | *n'*)) *pas*	(ne|n'=MOD(V)) pas=MOD(V)

3.2 (ne) plus

Rulename	Example sentence	Result of example	Pattern [12]	Action & Result
Non Plus	Sans compter non plus les drames humains et les vies gâchées	Non=MULT(plus)	(ˆ*pas*) *non plus*	Non=MULT(plus)
Ne V Plus	Mais elle ne travaillera plus toute seule dans son canton	ne=MOD(travaillera) plus=MOD(travaillera)	(*ne* | *n'*) (PRO | ADVˆ(*ne* | *n'*))* V (PRO | ADVˆ(*ne* | *n'*|*pas*|*jamais*))* *plus*	ne|n'=MOD(V) plus=MOD(V)
	Marie semble n' avoir plus aimé ce film	n'=MOD(avoir) plus=MOD(plus)		
(Ne) V Non Plus	Elle ne dit non plus mot	ne=MOD(dit) plus=MOD(dit)	(*ne* | *n'*)? (PRO | ADVˆ(*ne* | *n'*))* V (PRO | ADVˆ(*ne* | *n'*|*pas*|*jamais*))* *non plus*	ne|n'=MOD(V) plus=MOD(V)
	Mais ce qui est euh très dangereux c' est de n' avoir non plus son attention euh sur la route	n'=MOD(avoir) plus=MOD(avoir)		
Ne Plus Inf	...avait égalment clamé haut et fort son intention de ne plus venir sur le Tour de France	ne=MOD(venir) plus=MOD(venir)	(*ne* | *n'*) *plus* (PRO | ADVˆ(*ne* | *n'*))* V(ˆinf)	ne|n'=MOD(V) plus=MOD(V)

46

Non Plus Inf	Mais ce qui est euh très dangereux c'est de non plus avoir son attention euh sur la route	plus=MOD(avoir)	(^pas) (non) plus (PRO | ADV)* V(^inf)	ne|n'=MOD(V) plus=MOD(V)
Conj Non Plus NP	Nous avons affiché dans la colonne de droites les vies perdues et non plus les vies gachées	plus=MOD(vies)	(mais | et) non plus (mal (de | d') | beaucoup (de | d') | peu (de | d') | DET | PRP:det | (de | d' (^PRP)) | NUM | ADJ | ADV | PRO:DEM)* (NAM | NOM) | PRO	plus=MOD((NAM | NOM) | last PRO)
Conj Non Plus PP	Les commandes sont sur le levier de vitesse et non plus the orders are on the handle of speed and not more sur le volant	plus=MOD(sur)	(mais | et) non plus (ADV)* PRP	pas=MOD(PRP)
Conj Non Plus AdjP	La location d' appartement en proche banlieu est raisonable et non plus trop cher comme à Paris	plus =MOD(chers)	(mais | et) non plus (ADV)* ADJ	pas=MOD(ADJ)
Conj Non Plus AdvP	Il n y a pas tellement moins de pollution à Nancy qu' à Strasbourg et non plus beaucoup plus qu'a Epinal	pas=MOD(plus)	(mais | et) non plus (ADV)* ADV	pas=MOD(last ADV)

PP Non Plus	Sur le volant non plus	plus=MOD(sur)	PRP (N | PRO | DET | NUM | ADV(ˆ*pas*) | ADJ)* *non plus*	plus=MOD(PRP)
NP Non Plus	Les vies gâchées non plus	plus=MOD(vies)	(N | PRO) (DET | NUM | ADV(ˆ*pas*) | ADJ)* *non plus*	plus=MOD(N | PRO)
AdjP Non Plus	Trop cher non plus	plus=MOD(chers)	(ADJ) (ADV(ˆ*pas*)* *non plus*	plus=MOD(ADJ)
AdvP Non Plus	Beaucoup plus non plus	(second) plus=MOD((first) plus)	(ADV(ˆ*pas*) *non plus*	plus=MOD(ADV)

3.3   (ne) jamais

Rulename	Example sentence	Result of example	Pattern [12]	Action & Result
Que Jamais	Plus que jamais Virenque sera l'homme à abattre au sens propre comme au sens figuré	Jamais=COMP(que)	*que jamais*	jamais=COMP(que)
Ou Jamais[5]	C'est maintenant ou jamais	jamais=CC(ou)	POS-of-left-dependent-*ou* .* *ou jamais* .* ^POS-of-left-dependent-*ou*	jamais=CC(ou)
(Ne) V Jamais	Le premier présenté en 97 (n')a jamais été appliqué	ne=MOD(appliqué) jamais=MOD(appliqué)	(*ne* | *n'*)? (PRO | ADV^(*ne* | *n'*))* V(^pper) (PRO | ADV^(*ne* | *n'* | *pas*))* *jamais*	(ne|n'=MOD(V)) jamais=MOD(V)
	Dugarry rejette ces accusations et affirme ne s'être jamais dopé	Ne=MOD(être) jamais=MOD(être)		
(Ne) Jamais Inf	La sécurité routière demande à ses commerciaux à ses salaries de ne jamais utiliser le téléphone en conduisant	ne=MOD(utiliser) jamais=MOD(utiliser)	(*ne* | *n'*)? *jamais* (PRO | ADV^(*ne* | *n'*))* V(inf)	(ne|n'=MOD(V(inf))) jamais=MOD(V(inf))
Jamais (Ne) V	Le taux de réussite du bac général est même à un niveau jamais atteint 876 %	jamais=MOD(atteint)	*jamais* .* (*ne* | *n'*)? (PRO | ADV^(*ne* | *n'*))* V	jamais=MOD(V) (ne=MOD(V))

	Jamais un chef d'Etat étranger n'a été reçu avec autant de chaleur d'égards et d'intimité	né=MOD(a) jamais=MOD(a)		
Jamais NP	Jamais cette situation	jamais=MOD(soutien)	*jamais (mal (de | d')	beaucoup(de | d')	peu (de | d')	DET	NUM	ADV* ADJ	ADJ)* (PRO) N	PRO*	jamais=MOD(N|last PRO)
	Jamais ça	jamais=MOD(ça)		
Jamais PP	Jamais avant la fin de son mandat	Jamais=MOD(avant)	*jamais* (ADV)* PRP	jamais=MOD(PRP)
Jamais AdjP	Jamais trop chers	Jamais=MOD(chers)	*jamais* (ADV)* ADJ	jamais=MOD(last ADJ)
Jamais AdvP	Jamais très mal jamais=MOD(last ADV	jamais=MOD(mal)	*jamais* (ADV)* ADV	pas=MOD(last ADV)

4 Evaluation of the annotation scheme

To put the proposed annotation scheme for French negation into perspective it can be compared with the proposed dependencies from another annotation scheme. For this we will look at the French Treebank.

The French Treebank is a 1 million word corpus from the newspaper Le Monde (Abeillé & Barrier, 2004). Although this corpus is not really suitable for comparison since it deals with written language while our annotation scheme is developed for spoken language, it will be used anyway since there does not exist any corpus for French spoken language yet. The Treebank is annotated with parts of speech, inflectional morphology, compounds and lemmas, and syntactic constituent. Only a subset of the French Treebank is also annotated with functional tags. These tags, combined with information about the constituents have been transformed into dependency relations by the project EASY (Gendner et al., 2003). The dependency relations are automatically assigned and manually corrected.

Although it cannot be demonstrated with a value how much annotations from one annotation scheme are better than the ones from the other annotation scheme, it can be discussed how the annotations in both corpora relate to each other by noting their differences and similarities.

The annotation scheme of the dependency relations is not public so the dependency schemes cannot be compared directly. The comparison is therefore only based on some observations from the corpus which might not give a complete overview of all annotations that are applied to the negations under discussion.

4.1 Sentential Negation

In both schemes negative items relate to other items in the sentence by means of modifying dependency relations. The governors of the negations are however not always similar. Regarding sentential negation in simple sentences, the annotations are the same. Both *ne* and the second part *pas plus* or *jamais* are connected to the verb. The first big difference occurs in sentential negation of a sentence with auxiliary verb. Our annotation scheme proposes that negation is attached to the inflected verb of the sentence while the French Treebank attaches negation to the past participle. This is the case for sentential negation caused by all the negative words considered (*pas*, *plus* and *jamais*. *Jamais* is also consistent with this analysis when it occurs at the beginning of a sentence while its scope marker *ne* and past participle occur somewhere in the end.

An exception to this analysis can be seen when sentential negation follows a conjunction. Our annotations would lead to the proposed 'standard way' of annotation towards the verb while only the head of the verb phrase would connect to the conjunction (figure 3b). The annotations of negation in the

treebank however, are different from other cases of sentential negation when it follows a conjunction. The second part of the negation attaches to the conjunction (figure 3a):
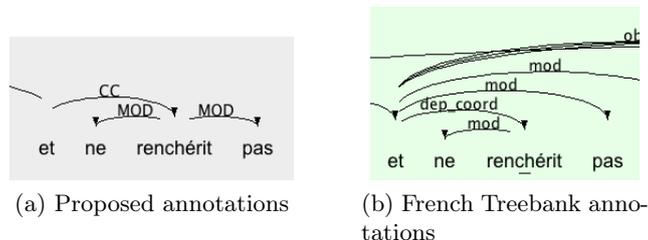

(a) Proposed annotations

(b) French Treebank annotations

Figure 3: *'and adds not'*

Another variation in the French Treebank is that sometimes, when there occurs sentential negation, the second part of the negation is attached to an element to the right which is not the past participle as is the standard case for sentential negation in this corpus. In figure 4b for example, *pas* causes sentential negation and is related to the element *beaucoup* at its right side. The annotations we proposed would as in all forms of sentential negation, be attached with a a dependency relation to the verb (figure 4a):

(a) Proposed annotations

(b) French Treebank annotations
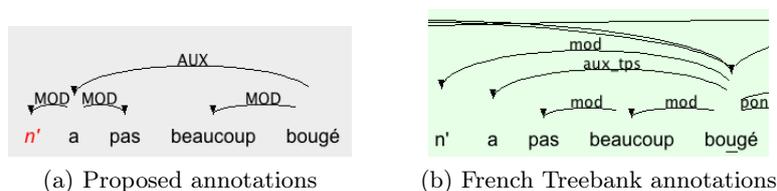
Figure 4: *'had not moved a lot'*

It is clear that there is not one consistent way to annotate sentential negation in the French Treebank, while we have proposed only one way of annotating sentential negation in the corpus of spoken French. Besides in the French Treebank, usually the items of negation are connected to the past participle although syntactically there is more support to attach them to the auxiliary verb as we have proposed in the annotation scheme (see section 2).

4.2 Clausal Negation

The items that express clausal negation are differently annotated than in the corpus of spoken French. *Ne pas* or *ne plus*, which occur before an infinitive are not two separate words but are written as one item with an underscore between the two words. The two words together have thus only one dependency relation (figure 5b). The governor does correspond to the one proposed in the corpus of spoken French (figure 5a). It is the infinitive

that follows. *Jamais* in the role of clausal negation has not been encountered in the French Treebank so it cannot be compared here.
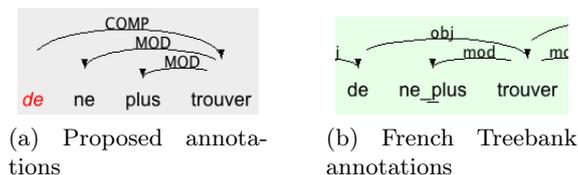

(a) Proposed annotations

(b) French Treebank annotations

Figure 5: *'not finding'*

4.3   Constituent Negation

Constituent negation occurs very rarely in the French Treebank. When it does occur it is almost always caused by ellipsis. In that case, the negation follows a conjunction. Sometimes the dependency relations that are added in this situation do not agree with the ones we proposed. It is proposed that *pas* in constituent negation is annotated to the head of the phrase it belongs to but in the French Treebank the negation is attached as a modifier to the conjunction:

(a) Proposed annotations

(b) French Treebank annotations
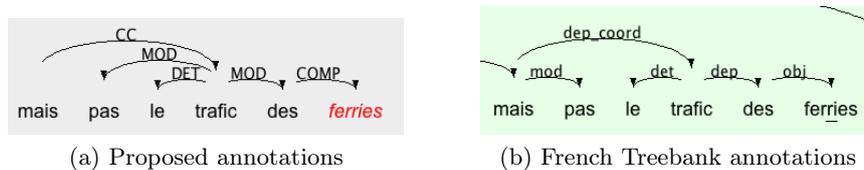
Figure 6: *'but not the ferry traffic'*

However, the Treebank contradicts itself with different annotations for the same structure, as can be seen in figure 7b. Here the annotation does agree with the proposed one.
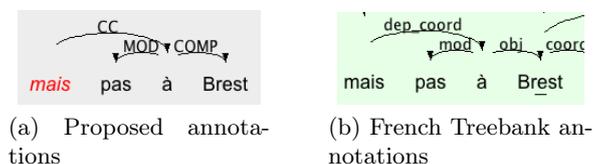

(a) Proposed annotations

(b) French Treebank annotations

Figure 7: *'but not to Brest'*

This is also the way constituent negation caused by *pas* or *non plus* (which is considered as one item: *non_plus*) is annotated when it is not caused by ellipsis, as seen in figure 8b. Constituent negation caused by *jamais* has not been found in the corpus.
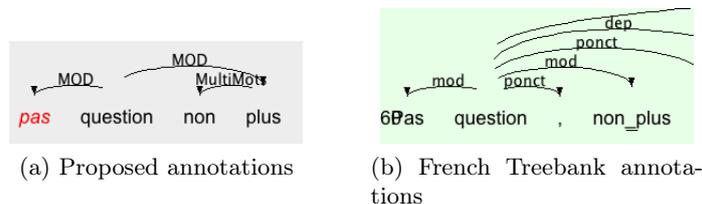
(a) Proposed annotations

(b) French Treebank annotations

Figure 8: *'no question either'*

Thus, in situations where constituent negation is not the consequence of ellipsis, the negations have the same governor, although non plus is annotated as one item in the Treebank and two items in the corpus of spoken French.

In short, in case of constituent negation, the proposed annotations only agree with the annotations of the French Treebank in restricted cases and that the annotations from the Treebank do not provide a consistent way to represent a specific type of negation (here constituent negation).

4.4 Multiword dependents

Items that are proposed as multiword dependents of negating items in the corpus for spoken French are differently annotated in the French Treebank. Some of them are already connected in the way they are written. They are considered as one item as can be seen in figure 8b. This is the case for *non_ plus*, *non_ plus* and *pas_ du_ tout*. They have together one dependency relation. This way of annotating is consistent with the way that is proposed. In the annotation scheme for spoken French there is also one dependency relation that connects the sequence of multiword expressions to another item in the sentence by means of one dependency relation. The difference lies only in the fact that the underscores in the Treebank are by multiword dependency relations in the spoken corpus.

An exception to this is the multiword expression *pas non plus*. Proposed is non=MULT(pas) and plus=MULT(non) and to attach *pas* to the item it negates. In the Treebank, *pas* has indeed as a governor the item that it negates but it has no connection with *non* and *pas*. These items are namely together (*non_ pas*) attached to the governor of *pas* as well.

4.5 Conclusion

There might be more variations in the dependency relations of negation in the French Treebank than are observed and mentioned here. Besides, some types of negation (for example constituent negation by *jamais*) do not occur in the French Treebank while they do in the corpus of spoken French. Nevertheless, it can be concluded that the annotation scheme for negation proposed for the corpus of spoken French, differs on quite some aspects from

the dependency relations for negation used in the French Treebank. Our annotation scheme is however based on syntactic reasons for which differences with the French Treebank are no reason to adapt the proposals. Instead, the question arises how much the dependencies in the French Treebank are linguistically motivated. An insight into how its annotation scheme is developed could clarify this. Further, it is noted that the dependencies in our scheme are generally more consistent then the Treebank ones. Not only does the written corpus have for equal structures different dependency relations, there is also no uniform way to handle a specific type of negation. Sentential negation for example, can be annotated with many different governors in the written corpus, while the proposed annotations only point to one type of governor.

5 Evaluation of the automatic implementation

In section 3 rules are presented to automatically implement the proposed dependency relations for negation. Here it will be evaluated how well these rules actually cover the proposed annotations. First a methodology is presented and afterwards the results and errors will be discussed.

5.1 Methodology

In order to test the correctness of the automatic implementation of the annotation scheme, its generated annotations will be compared with so-called 'golden' annotations that meet all the prescriptions from the annotation scheme. The 'golden' annotations consist of manual annotations that are assigned according to the guidelines of the scheme. Because there are no problems or contradictions encountered while implementing the manual annotations, all annotations meet the prescriptions from the scheme and can indeed function as the golden standard.

The manual annotation task is carried out by the developer of the annotation scheme on the same corpus that was used for the design of the scheme. Different annotators and corpora for the development of the scheme and the implementation of the annotations would have given a better insight into the all-round coverage and clearness of the scheme. But unfortunately, a lack of annotators and corpora prevented this possibility. The lack of corpora is due to the fact that there is only one corpus of spoken French that is already annotated with the annotation scheme from Cerisara, Gardent and Anderson (2010). Annotations from corpora that use a different annotation scheme might conflict with the developed scheme for negations while corpora without annotations at all will need to be parsed anyway after the treatment of negations. This might also cause conflictions. Therefore we will only look at the corpus developed by Cerisara, Gardent and Anderson (2010). Because this corpus is quite small (50,000 words), and restricted to a specific type of spoken French (radio-broadcasts) the results of the evaluation are not claimed to be representative for spoken French in general.

The corpus used for the evaluation is already annotated with dependency relations from an existing annotation scheme (see section 1 for details). The proposed annotations for negation thus actually consist of manual corrections of the original dependency relations. Which item receives an annotation in this way and which one not, can be deduced from the guidelines in the proposed scheme. For example, *pas* is only annotated when it performs verbal or constituent negation, or when it is the complement of *ou*. Other occurrences of *pas*, as disfluency or noun for example, are not handled in the annotation guide and are therefore not manually annotated. Another example is *non* that is only annotated when it combines with *pas* or *plus*.

After the golden annotations, the corpus is annotated with the automatic

dependency relations. For all instances of *pas* for example, it will be checked what kind of POS tag it carries and in what sequence of word categories it occurs. On the basis of that it is determined whether and what kind of annotation the item receives. Because the automatic rules consider every occurrence of a certain word, the evaluation is also performed on every occurrence of a word for which an annotation is proposed. (This includes all instances of *pas*, *plus*, *jamais*, their scope markers and their multiword dependents.) There is no negation detector to evaluate only the negative uses of the items.

In what extend the automatically assigned annotations are correct or not can now be estimated by comparing the automatic annotations with the golden annotations: It is considered that any automatically assigned dependency relation (labelled with PROPPAS) is correct if and only if there exists a gold dependency (labelled with GOLDPAS) with the very same dependent and governor nodes. This means that if an item is assigned a dependency relation with a different governor than the golden annotation proposes, it will be counted as an error:
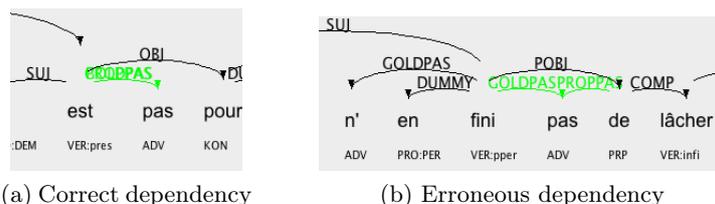

(a) Correct dependency   (b) Erroneous dependency

Figure 9

Another type of error occurs when an item is automatically annotated but not manually, or the other way around:

(a) Automatic annotation without golden (last *non*)   (b) Golden annotations without automatic (*n'* and *pas*)
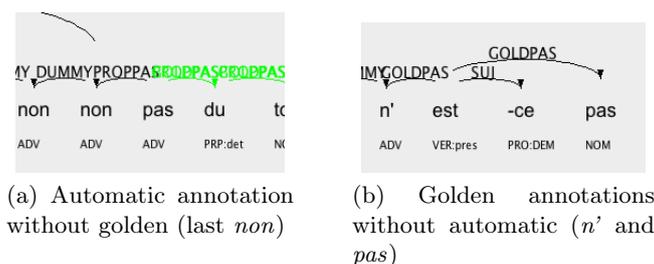
Figure 10

When an item is not manually annotated and neither automatically there is no error.

Widely used evaluation metrics for detection problems are precision and recall. They also have been used for the evaluation of dependency annotations (Lin, 2003). However, there is also a metric that is specifically used

for the evaluation of dependency annotation tasks: the *LAS score* (labelled attachment score). This metric has been used to evaluate the original annotations of the corpus (Cerisara, Gardent and Anderson, 2010) and is the official scoring metric of the CoNLL parsing evaluation campaigns. Therefore, the LAS metric will be used here as well such that the scores can easily be compared with other projects. The LAS score is defined as the percentage of tokens that is attached to the correct governor with the correct dependency label. This means that it will only consider annotations like in figure 9a correct and counts annotations such as of figure 9b, 10a and 10b as errors.

The rate of the errors of the type presented in figures 10a and 10b are also estimated separately by the metrics *insertion* and *omission*. The insertion value indicates the percentage of items that is assigned an automatically assigned dependency relation but should not have one, while omission is the percentage of items that does not have an automatically assigned dependency relation but should have one.

5.2   Results

As seen in section 3, the automatic implementation of the annotation scheme consists of three parts containing multiple hierarchal rules. There is a part that deals with the annotations that are involved with *pas*, another for *plus* and one for *jamais*. Because the annotations proposed in one of the three parts do not depend on the annotations in another part, evaluation values can be obtained for the automatic implementation of each part separately. This will provide an insight in how the evaluation score of the entire annotation scheme is build up. In the following paragraphs the evaluation scores for each part belonging one of the three negating words are presented separately. After that the evaluation score for the all-round automatic implementation of the annotation scheme will be given.

5.2.1   (Ne) Pas

The first part of the annotation scheme considers the negative item *pas* including its scope marker and multiword dependents. The automatic implementation of all the items together that are handled in this part have a LAS score of 0.967 with a confidence interval of 0.013. Meaning that 96,7% of the automatically generated annotations are correct. The score for the item *pas* alone is 0.958. After *pas*, *ne* is the most frequently annotated item (274 times), and has a LAS score of 0,981. Other words that are used in combination with pas and therefore annotated, occur considerably less frequent (below 10 times). This makes that the dependency relation for one occurrence of such an item can already have a big influence on its evaluation score. The LAS value for *non* is with one error the lowest: 0,833, while *du*, *tout* and *plus* are always correctly annotated leading to a score of 1. All

items together give an all-round LAS score of 0,967. The insertion and emission ration are quite low (rel. 0,007 and 0.011), which means that there are very few items annotated while they should not, and items not annotated while they should.

	Total	Pas	Ne	Non	Plus	Du	Tout
LAS score	0.967	0.958	0.982	0.833	1.0	1.0	1.0
Nr. gold occurrences	704	404	274	6	2	9	1
Confidence interval	0.013						

(a) LAS scores

Insertion ratio	Omission ratio
0.007	0.011

(b) Insertion/omission ratio

Table 1: Results automatic implementation *pas*

To get an ever more detailed view of how the different rules of the automatic implementation contribute to the overall result, we can not only make a distinction between the sections used to annotate one of the negating elements (*pas*, *plus*, and *jamais*), but also between the different rules that are used within one of these parts. For *pas* for example, there are 11 different rules to handle all its occurrences. The influence of each of these 11 rules is represented in figure 11. It can easily be seen how the addition of a certain rule contributes to the overall evaluation scores.

The graph shows that words as *du* and *tout* are treated with only one rule, while *pas* has its annotations due to all 10 of the rules. Besides, it is easily notable that the total LAS score for the pas part of the evaluation scheme is mostly build up from the rule (NE) V Pas and (Ne) Pas Inf which deal with verbal negation. Other rules improve the evaluation score as well, but on a much smaller scale.

The head of NP to the left of Non Plus, is difficult to find. Better to do it with already existing annotations.
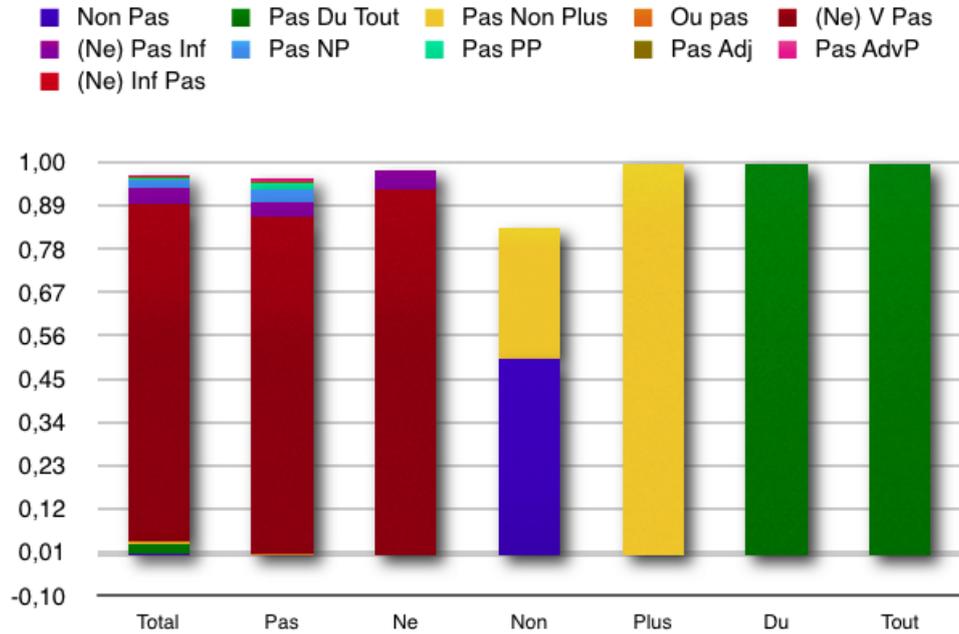
Figure 11: Contributions to the LAS scores for *pas*

5.2.2 (Ne) Plus

The negative *plus* occurs a lot less frequent than *pas*, as can be seen in table below. It contribution to the scores of the entire annotation scheme are therefore a lot smaller than annotation for *pas*. The annotations added to handle *plus* add dependency relations to 3 different items. *Plus*, *ne* and *non*. In the table below the overall score as well as the LAS score of the separated items can be seen. The confidence interval for the overall score is however small due to the low frequency of the annotated items.

Graph 12 shows how much each rule contributes to the overall scores on top of the former rule.

	Total	Plus	Ne	Non
LAS score	0.842	0.75	0.933	1.0
Number of gold occurrences	38	20	15	3
Confidence interval	0.11			

(a) LAS scores

Insertion ratio	Omission ratio
0.0	0.13

(b) Insertion/omission ratio

Table 2: Results automatic implementation *plus*

The omission ration indicates that all items that are manually treated are also automatically treated. The other way around is not the case though. Errors are thus only of the type automatically annotated while this should not be the case, or annotated with a dependency relation with a wrong item.
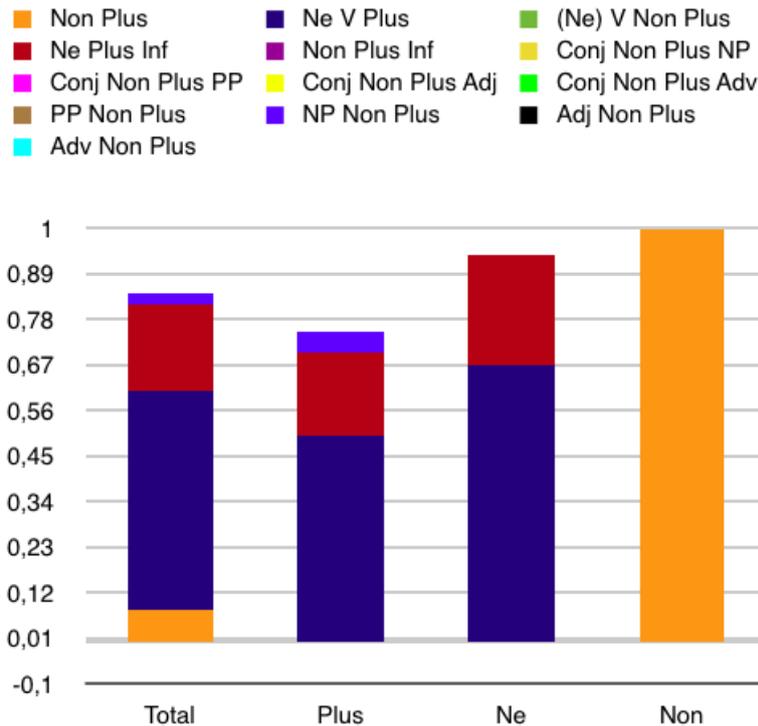

Figure 12: Contributions to the LAS scores for *plus*

At graph 12 it can be seen that lots of rules were not of hand in the corpus. This might be due to the little corpus and small amount *plus*. In other, larger corpora the rules might be useful. However, it is not tested here how well these rules work adequately since there are no (unbiased) instances to test them on.

5.2.3   (Ne) Jamais

Jamais occurs just as little in the corpus as the negative use of *plus* but its LAS score is overall better. In the table below it can be seen that 97,5% of the annotations were assigned correctly. Besides, its confidence interval is lower than that for *plus* which means that the reliability of its results is bigger. However, it is still a high interval in comparison with that of *pas*. This is probably due to its low frequency. The omission value shows that

	Total	Jamais	Ne
LAS score	0.975	0.96	1.0
Number of gold occurrences	40	25	15
Confidence interval	0.048		

(a) LAS scores

Insertion ratio	Omission ratio
0.025	0.0

(b) Insertion/omission ratio

Table 3: Results automatic implementation *jamais*

there is not one dependency relation assigned while they should not have been assigned. A small percentage of the annotations are not assigned an evaluation while they should have one. Graph 13 demonstrates that there
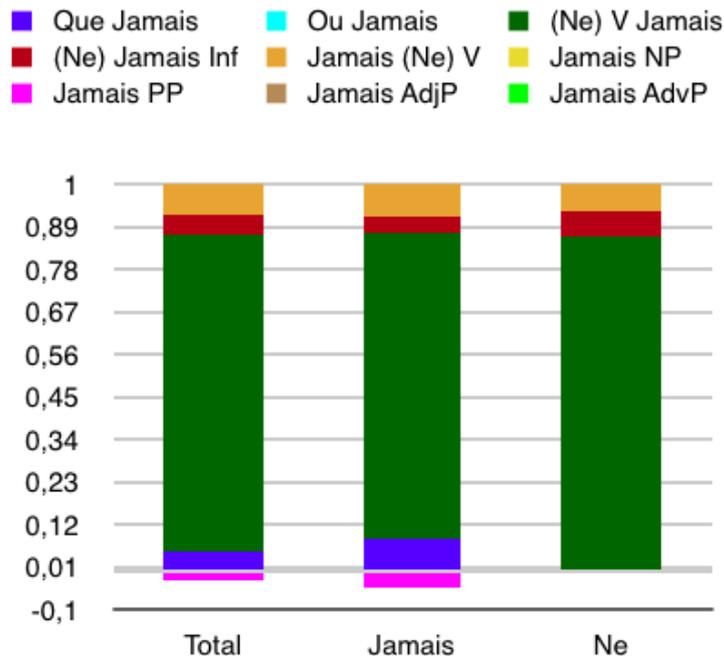

Figure 13: Contributions to the LAS scores for *jamais*

are 4 rules that actually contribute to the annotations in this corpus. Other rules were not applied here. Further, it can be seen that there is 1 rule which gives a negative score (Jamais PP). This however, does not necessarily mean that it is not working correctly, it can also be the case for example that *jamais* is a disfluency and should be left unannotated.

5.2.4 Results and overall errors

	Total	Pas	Plus	Jamais	Ne	Non	Du	Tout
LAS score	0.961	0.958	0.773	0.96	0.980	0.889	1.0	1.0
Nr. gold occurrences	782	404	22	25	304	9	9	9
Confidence interval	0.024							

(a) LAS scores

Insertion ratio	Omission ratio
0.008	0.017

(b) Insertion/omission ratio

Table 4: Results automatic implementation overall

Although the LAS scores for all annotated items are pretty high, some items have better scores than others. However, these scores cannot be considered without taking into account the frequency of the items. It can be seen that *plus*, *jamais*, *non*, *du* and *tout* occur very little in the corpus. The annotation of one occurrence of these items can thereby influence the score for this item a lot. Besides, the range of the occurrences of these items is not large enough to state whether or not the rules work appropriately. The only thing that can be stated is that in this small corpus they work well on the few instances of these items.

The implementations for *ne* and *pas* however, are applied to a lot more occurrences. It has lead to a lot of correspondences between the golden annotations and the automatically assigned ones. Because these items occur frequently in the corpus (in comparison to the other items), it is likely that the automatic implementations for these annotations will indeed work well outside this corpus as well.

The insertion and omission ration of the annotations are quite low which means that there are few items assigned a dependency relation while they should not have one, and the other way around. Errors are mostly due to wrong dependencies to wrong governors.

The unexpected errors of the automatic implementations can be divided in three types:

First there are errors which are due to wrong POS tagging. Next to the obvious errors of POS tagging such as confusing the adverbial *pas* with the noun *pas*, there are also problems due to more subtle tagging issues such as when *de* is annotated as preposition but can actually be considered a determiner. More errors could be avoided by assigning one POS tag to multiple items when appropriate. Such as *sans doute* which could be assigned as an adverb but which is now has an adverbial function but should now be

recognized by a preposition and a noun.

Transcriptions that are not what expected have lead to problems as well. For example, *peut-être* is transcribed as two words while it could be transcribed as one.

Another source of problems comes from the lack of correct sentence boundaries. If it would be clear where one sentence ends and another starts dependencies between two different sentences could be avoided.

Other errors than the ones mentioned so far are expected and cannot be avoided with the proposed annotation scheme. One of them is due to the ambiguity of the positive and negative use of *plus*, and the other one is caused by negative items that are disfluencies.

In only one case, an automatic rule is not working in the way it should. This programming of this rule should be verified.

Another corpus must be checked to validate the general applicability of the annotation scheme and its automatic implementations.

It can be concluded that the automatic implementation of the annotations obtain a LAS score of 0.96% on the corpus on which they where developed. In order to say anything about the general applicability of the rules more data should be considered.

6   Conclusion

This project has proposed dependency relations for negation. They are based on linguistic characteristics of the negations, notational issues of dependency grammar and earlier decisions in the annotation scheme. In this way we aimed to present annotations that are acceptable from different perspectives.

We tried to deal with as much occurrences of the negative items *pas*, *plus* and *jamais* as possible. For this, examples in and outside the corpus where analyzed. The different meanings, types and constructions of the negative items were described and corresponding annotations were proposed.

Manual annotations of the corpus of French radio broadcast that are based on the proposed annotation scheme could handle all instances in the corpus. The automatic implementations of the annotation scheme have quite good scores in comparison to the manual ones: they have a LAS score of 0.961 with a confidence level of 0.024. However, for the manual annotations as well as for the automatic implementations biases cannot be excluded because the same corpus is used for the development as the implementations of the annotations. Another issue that should be taken into account is that the results are achieved by means of a very small corpus. They do not give by any means a representation of the performance in general.

The annotation scheme can be applied to the corpus for which it was developed, the Treebank of French radio broadcast, and can be implemented next to the already existing annotations. At the same time the automatic implementation of the rules can be checked further when the Treebank will be augmented.

References

[1] A. Abeillé and N. Barrier. Enriching a french treebank. In *Proceedings of Language Ressources and Evaluation Conference*. Citeseer, 2004.

[2] A. Abeillé, O. Bonami, D. Godard, and J. Tseng. The syntax of french de-n'phrases. In *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar. Stanford (Cal.): CSLI Publications*, pages 6–26, 2004.

[3] C. Gardent, C. Cerisara, and C. Anderson. Building and exploiting a dependency treebank for french radio broadcast. 2010.

[4] V. Gendner, G. Illouz, M. Jardino, P. Paroubek, L. Monceaux, I. Robba, and A. Vilnat. Proposition de protocole d,Äôévaluation des analyseurs syntaxiques du français: Peas. *TALN 2003*, 2003.

[5] D. Godard. French negative dependency. *Handbook of French Semantics, Stanford, CA: CSLI Publications*, 2004.

[6] M.A. Jones. *Foundations of French syntax*. Cambridge Univ Pr, 1996.

[7] D. Lin. Dependency-based evaluation of minipar. *Treebanks: building and using parsed corpora*, pages 317–332, 2003.

[8] P. Rowlett. *Sentential negation in French*. Oxford University Press, USA, 1998.

[9] L. Tesnière. *Eléments de syntaxe structurale*. Libraire C. Klincksieck, 1959.