

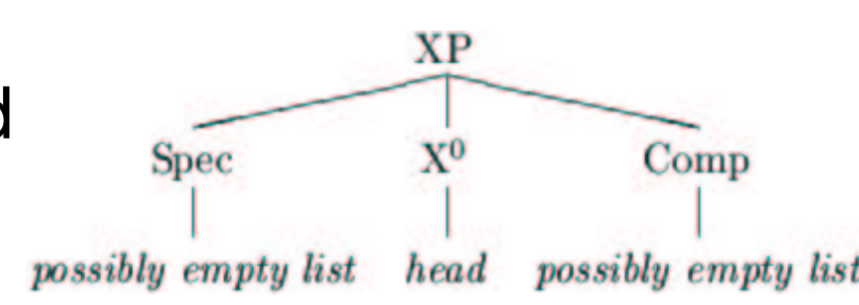
The FreeText Project

- French in Context: An advanced hypermedia CALL system featuring NLP tools for a smart treatment of authentic documents and (relatively) free production exercises. <http://www.latl.unige.ch/freetext/>
- Target: Intermediate to advanced learners of French
- communicative approach to second language acquisition, focused on learners' inputs
- NLP Tools for intelligent feedback:
 - Sentence parser
 - Sentence viewer (color grammar)
 - Diagnosis tool
 - Speech synthesis of software and learners' sentences
 - Sentence reformulation tool
- funded under the User-friendly Information Society (IST) program of the 5th framework program of the European Commission and by the Swiss Federal Office for Education and Science
- Four partners participate in the FreeText project:
 - University of Manchester Institute of Science and Technology, Department of Language Engineering (UK)
 - University of Geneva, Department of Linguistics (CH)
 - Université Catholique de Louvain, Centre for English Corpus Linguistics (B)
 - Softissimo SARL, Paris (F)

FIPS Parser

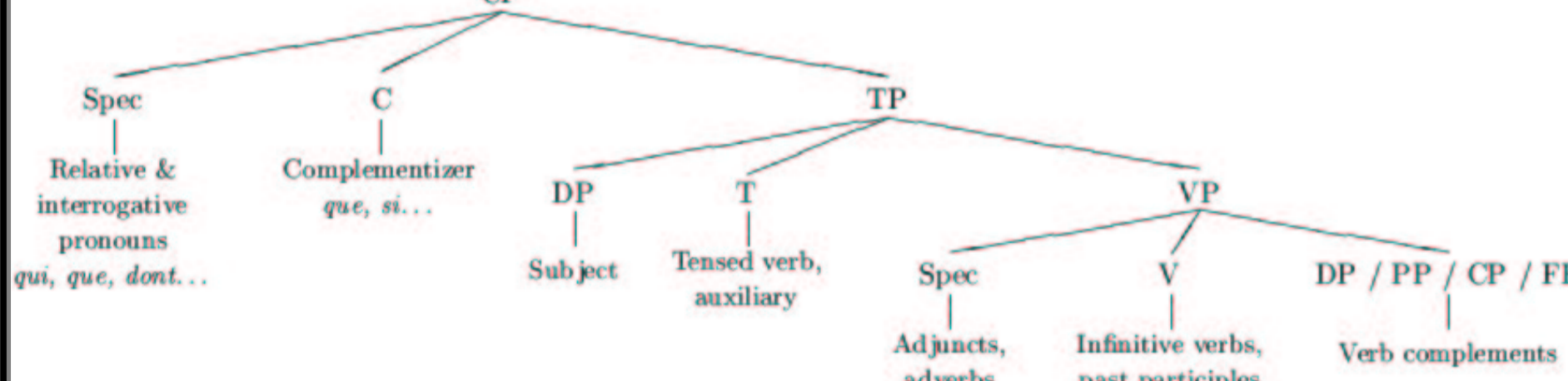
FIPS is a large-scale, in-depth, modular syntactic parser based on Chomsky's *Principles & Parameters* Theory, developed since 1990 at the Laboratory For Language Analysis and Technology (LATL). It can be tested at <http://www.latl.unige.ch/>

FIPS output structures are based on a simplified version of the X-bar schema. Spec and Comp are two lists of projections that can be combined to the left and right of their mother projection



A maximal projection XP corresponds to each lexical (A, Adv, N, V) or functional category (D(eterminer), T(ense), F(unction)). The position X⁰ is called the head.

Here is FIPS canonical structure for a French sentence:

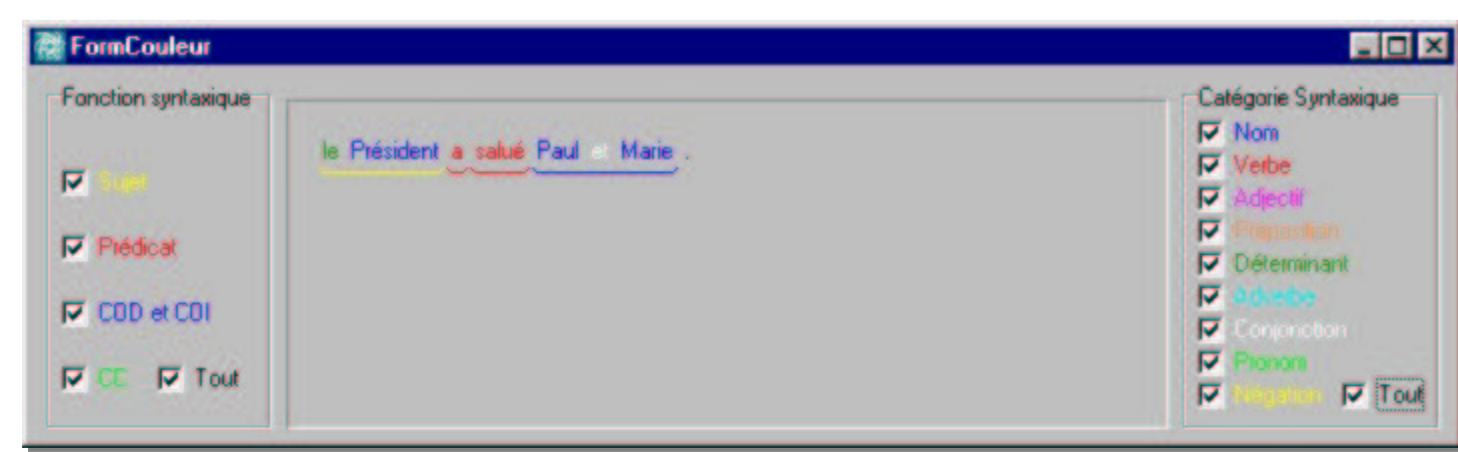


The DP (determiner phrase) is usually described as NP in other formalisms, and includes the NP in its complements list.

Here is the structure for "Le président a salué Paul et Marie":

[_{TP} [_{DP} le [_{NP} président]]] [_T a [_{VP} salué [_{DP} [_{ConjP} [_{DP} Paul] [_{Conj} et [_{DP} Marie]]]]]].

These structures are rather complex and hard to understand for learners. Therefore we have developed a color representation of the structure and a simplified set of grammatical labels to be used in CALL software.



Ph[GN][Déf le] [N président]][GV[Aux a] [PartPassé salué] [GN [GN Paul] [Coord et] [GN Marie]]]].

These labels will be used in syntactic trees like the example above.

We are currently developing a new XML output for the parser, from which further output can be derived. See the box in the lower right corner.

Syntactic and "semantic" error detection

Sébastien L'HAIRE, Anne VANDEVENTER FALTIN, University of Geneva, Switzerland

Presented at EUROCALL 200Nijmegen, August 29th – September 1st

Error detection:

The learner answers an exercise. If this answer **doesn't exactly match the answer stored in the system**, it will be submitted to a 3-step procedure:

- 1) **Spell checker**. If the answer still doesn't match, then

GBGen: Sentence Generator

GBGEN is a deterministic, large-scale sentence generator based on Chomsky's P&P Theory. Sentences are generated as deep structures, some movements are applied to obtain a surface structure and finally some morphological operations are performed. No backtracking or pseudo-parallelism is allowed.

GBGEN's input is hybrid interface structures called Pseudo-semantic structures (PSS). These structures contain both abstract semantic information and lexical items.

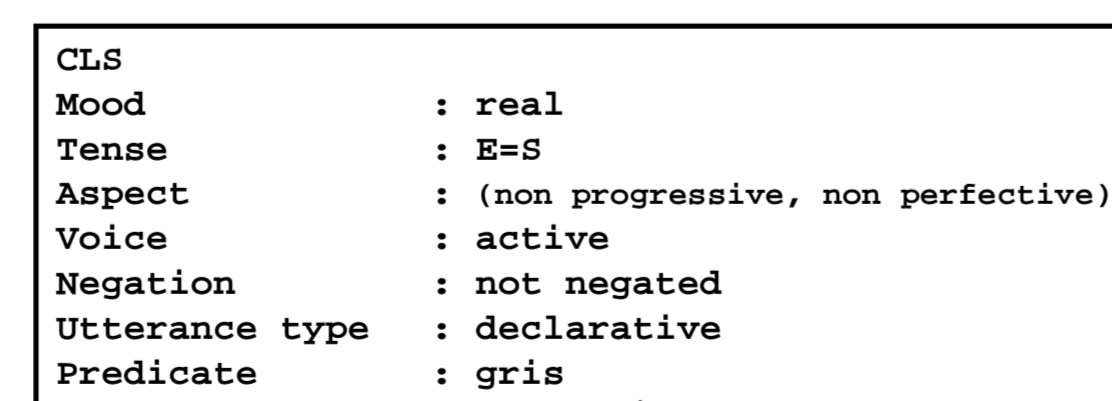
PSS are divided into 3 subtypes: CLS (clause structures) which correspond to states and events and contain verbs and adjectives, DPS (DP-Structures) which correspond to nominal structures and contain nouns, and CHS (Characteristic structures), which correspond to phrase modifiers and contain adjectives and adverbs.

PSS can combine into hierarchical structures; they contain an unordered list of other PSS, called *satellites*. See the **example structures in 3**. PSS are automatically derived from FIPS output structures.

GBGEN can be used for machine translation, 'semantic' error detection, sentence reformulation and other operations.

3) 'Semantic' checking If the analysis is complete

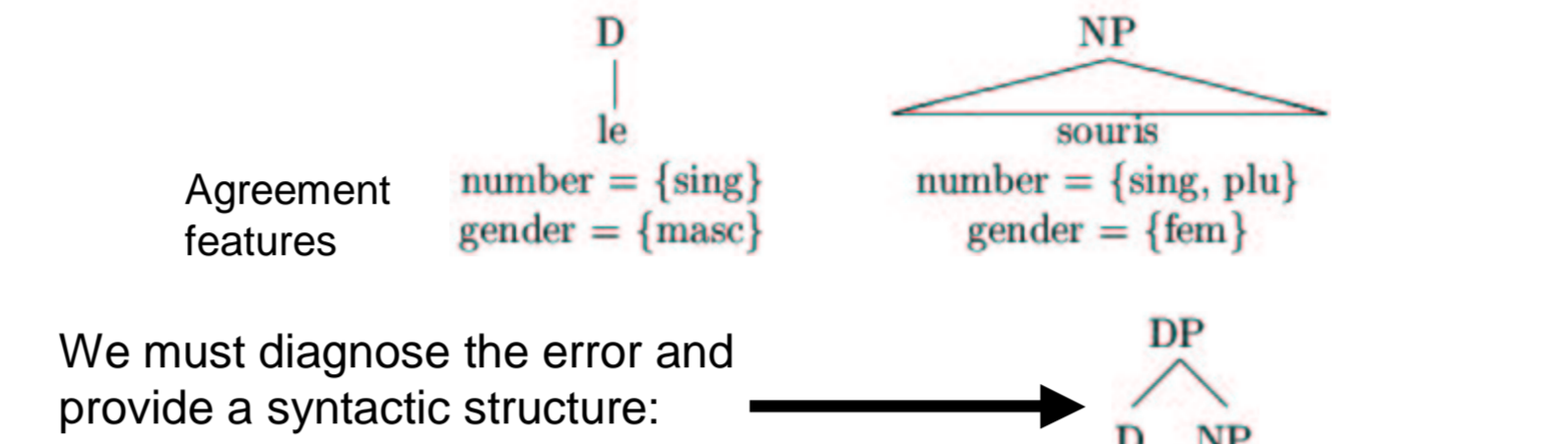
The learner sees a picture with a small, grey mouse. The question is: «De quelle couleur est la petite souris?» (*Which color is the small mouse?*). The expected answer stored in the system, which has to be as complete as possible, is «La petite souris est grise» (*The small mouse is grey*). Here is a simplified PSS structure (see above) for this expected answer (arrows denote the relation 'is satellite of'):



The CLS contains the lexical item «gris», which is a state. Verbal properties are given by the abstract values. From this structure, the system would generate the verb être (to be), in the tense given by the values, with the attribute «gris» (grey). The single satellite of the CLS, a DPS, contains the experiencer of the clause (in this case: the subject). It contains the noun «souris». The determiner is represented by the abstract operator «the», which would generate a definite determiner «le». This DPS also has a satellite, the adjective «petite» (small), which is a subset of a class (in this case: the class of mice), like all the adjectives in CHS.

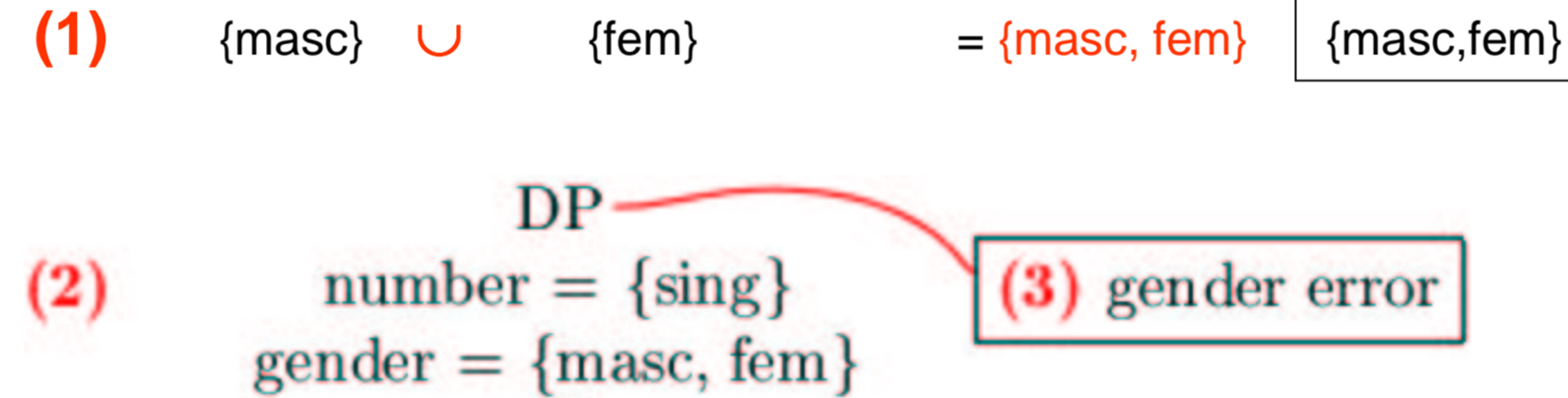
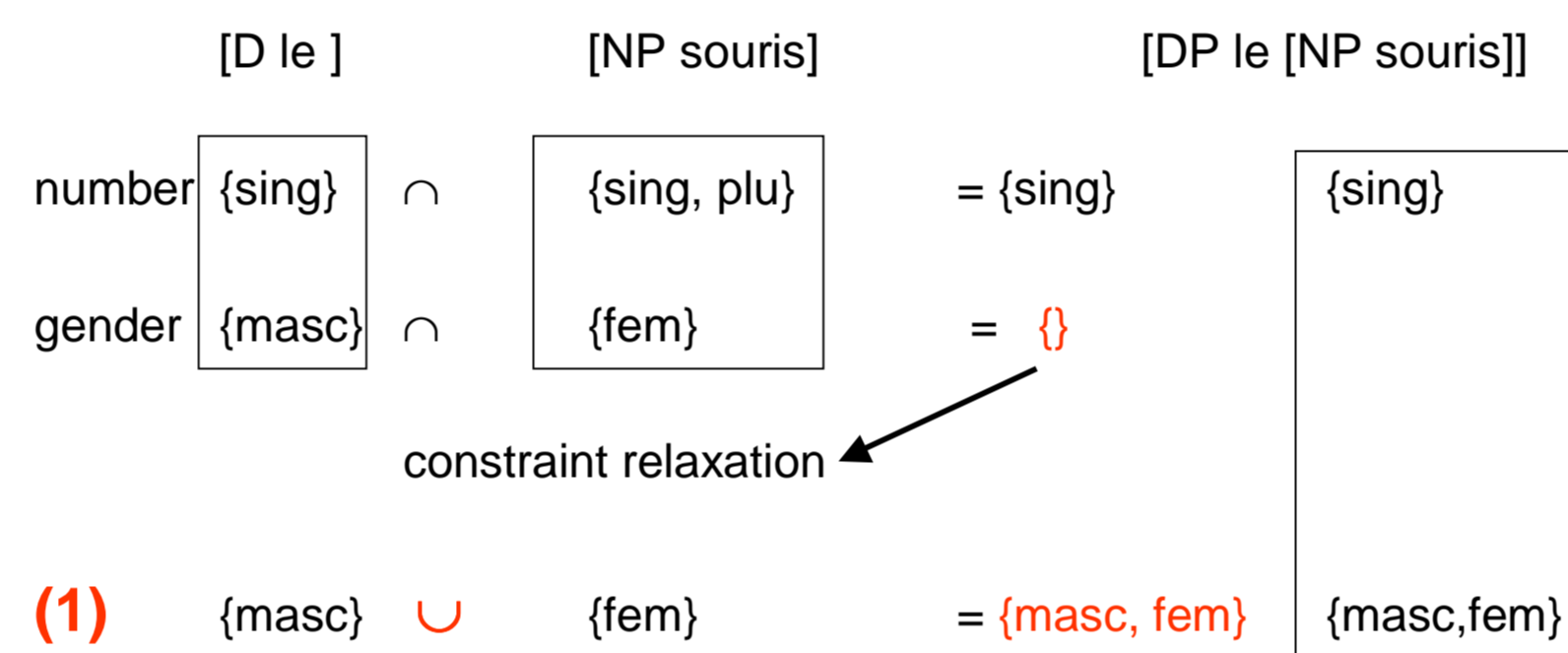
2.A) Constraint relaxation

The learner types in: le souris
the-masc-sing mouse-mice
Result of parsing: no complete analysis but chunks:

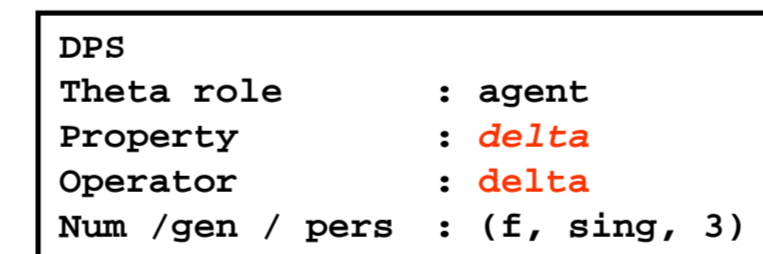


We must diagnose the error and provide a syntactic structure:

Agreement constraint:
•To combine into a full DP, D and NP must agree in number and gender
•The intersection of their feature sets must not be empty
number(D) ∩ number(NP) ≠ {}
AND
gender(D) ∩ gender(NP) ≠ {}



Let's assume that the learner answered «elle est grise» (*she is grey*). The learner's answer CLS is the same. As CLS satellite, we have the following DPS, with no CHS attached to:



delta means that the property or the value is left empty.

Both PSS structures are compared. Then the system can determine if both structures are compatible. In this case it depends on whether pronominalization is allowed. If not, we will flag the output structure and give the learner a relevant feedback.

Assuming that we accept elliptic answers, we can also consider correct «la souris est grise» and «grise», but we have to reject «gris», which can be parsed but does not have the correct gender.

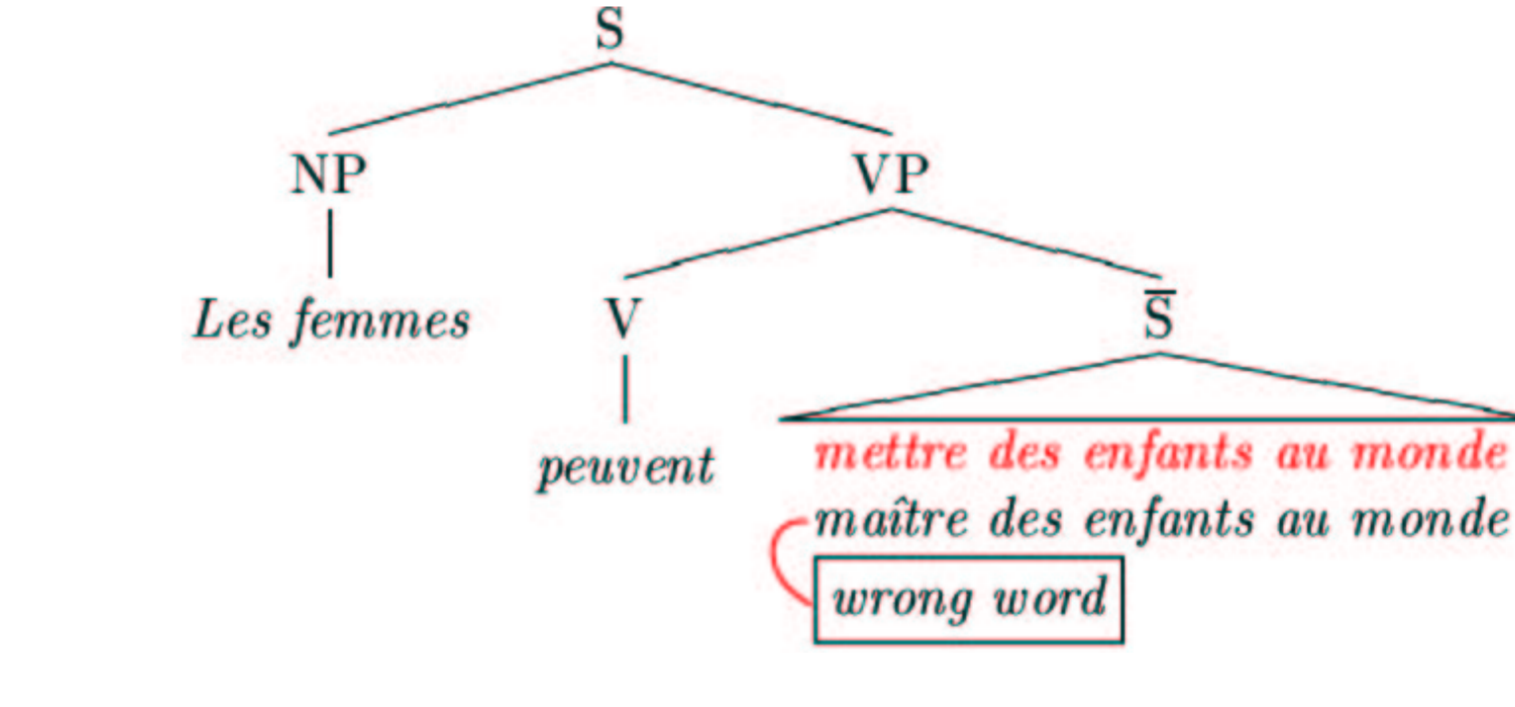
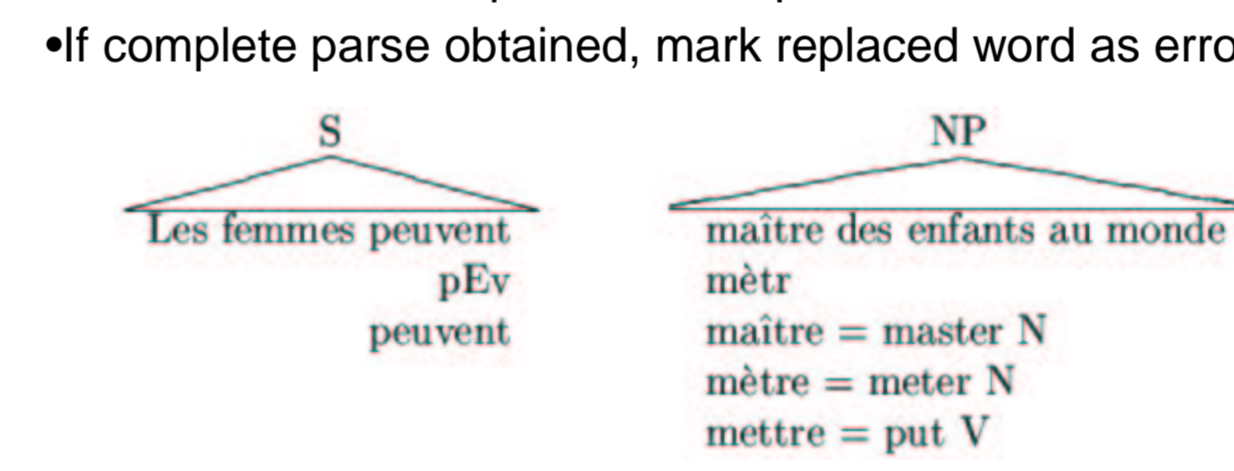
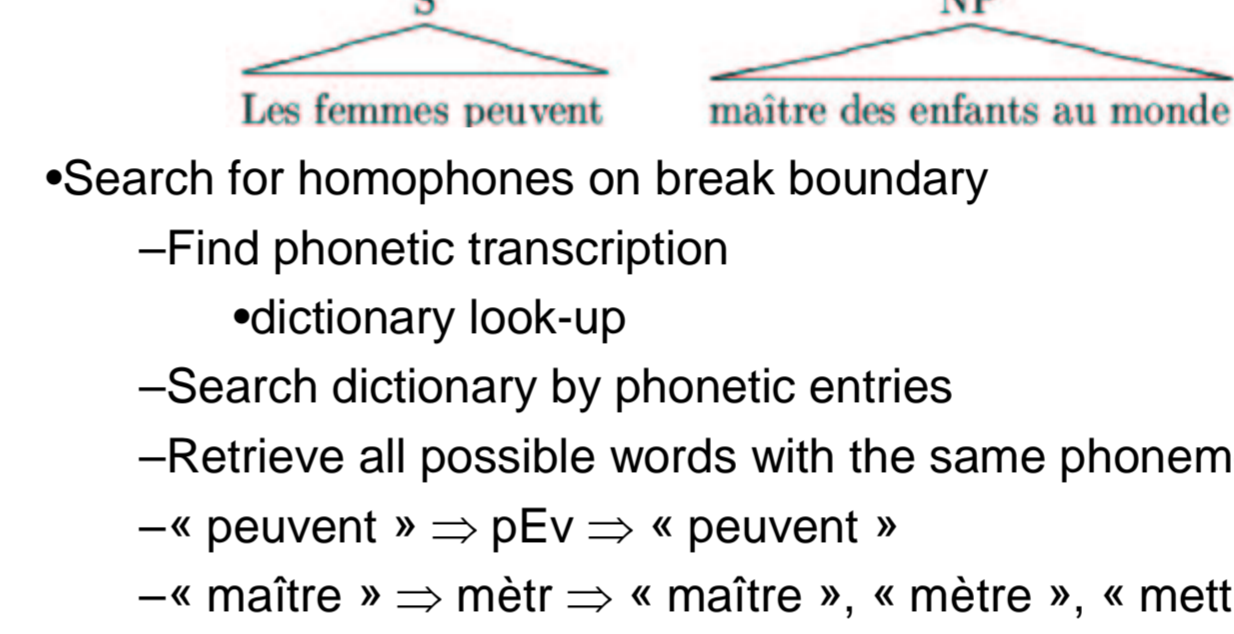
The system will be able to discover discrepancies between the learner's answer and the answer stored in the software by comparing the relevant PSS.

If the question in a pronominalization exercise is: «as-tu vu les voitures?» (*did you see the cars?*), the correct answer will be «oui je les ai vues». The learner can answer «oui, je les ai vus», which is syntactically correct if the pronoun «les» is masculine plural. However, it is wrong: the pronoun replaces «les voitures» (feminine plural). Therefore, there is an **agreement error** between the pronoun and the past participle of the verb «voir» (to see) which must be the feminine plural form *vues*. The answer is **semantically wrong** within the context of the exercise.

2) Grammar checker If no complete analysis:

2.B) Phonetic reinterpretation

The learner types in: les femmes peuvent maître des enfants au monde
The women can master-noun a-plural children in the world
Result of parsing: no complete analysis but chunks:



So far, we plan to use this technique with **replacement exercises** like *pronominalization, active / passive constructions, nominal sentences / complete sentences, infinitive / imperative* etc. Some tests may also be applicable to open questions.

We also plan to use a **synonym dictionary**, in order to accept synonyms, hyponyms or hyperonyms. If we ask «Qu'est-ce que Jean a acheté?» (*What did John buy?*) and expect «Jean a acheté un perroquet» (*John has bought a parrot*), we would like to accept «Jean a acheté un oiseau» (*John has bought a bird*), which contains the hyperonym (or generic term) «oiseau» for the expected «perroquet».

2.C) Chunk reinterpretation

Idea:

- If a sentence is ungrammatical, it will result in partial analyses
- in examining these partial analyses, one should be able to determine why the chunks could not combine
- series of specific rules for specific cases

Can be used as complementary information to confirm another diagnosis

1) Could be done by phonetic reinterpretation

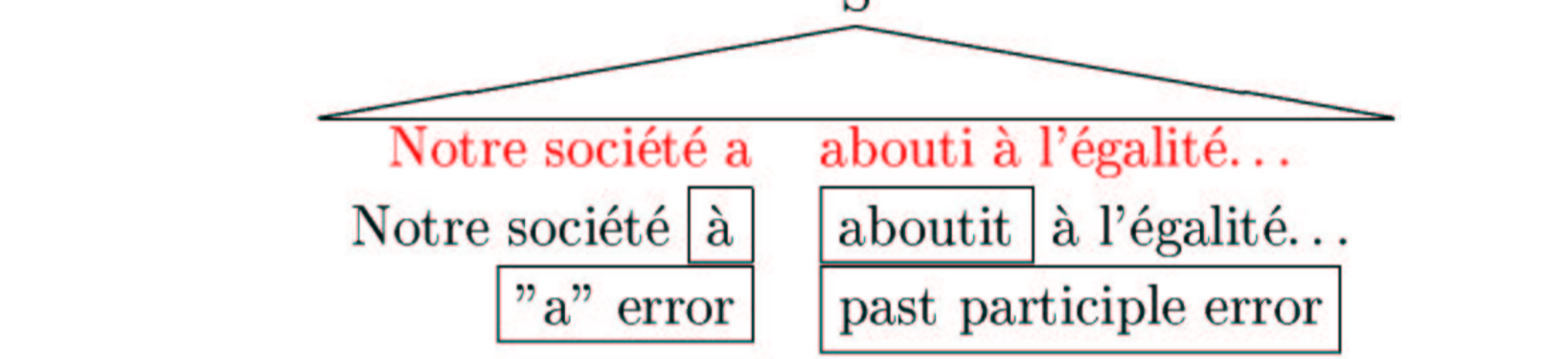
- The learner types in: Notre société à aboutit à l'égalité des hommes et des femmes
- Our society to succeeds to the equality of-the men and of-the women



Right-most item of left chunk is «préposition à».

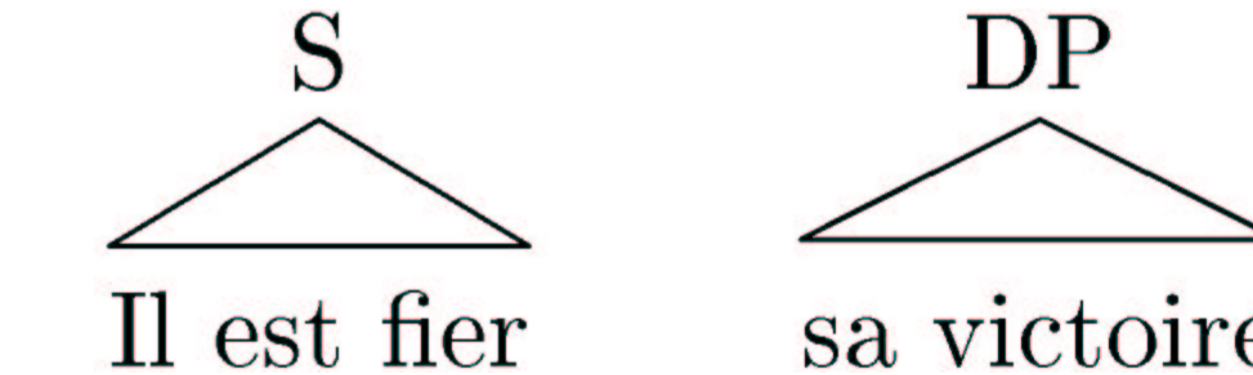
Left-most item of right chunk is «non-infinitive verb»

Replace «à» by «a» and verb in present tense by past participle and parse again.



2) Could be done by constraint relaxation

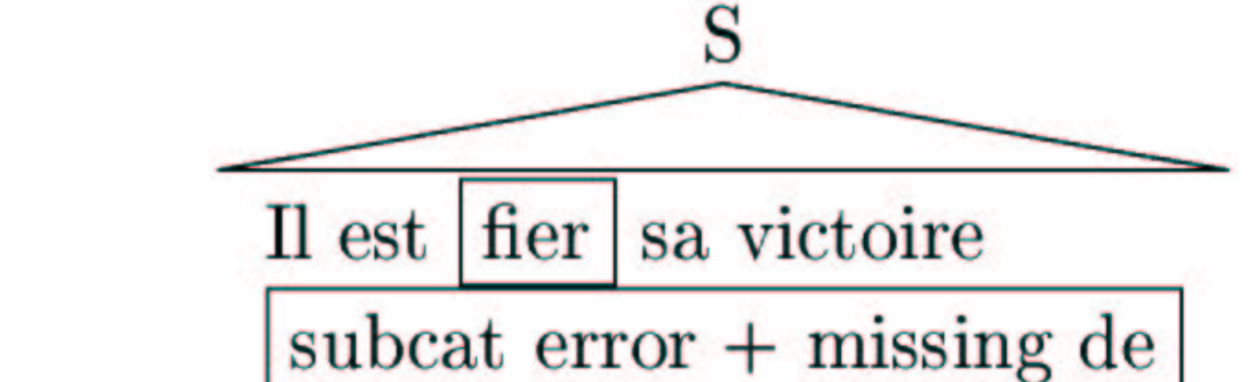
The learner types in: Il est fier sa victoire.
he is proud his victory



Right-most item of left chunk is «adjective fier» AND takes a PP complement

Right chunk is «DP»

Signal missing «de» preposition and subcategorization error



XML error tagging

FIPS output is a XML document.

Let us consider the sentence "les chats dort", where the plural subject does not agree with the singular verb.

The whole document is included in <LATLPARSE>.

A complete sentence or a chunk of analysis is written in <ANALYSIS>. <PROJ> contains projections and <HEAD> the projections' head. <BAR> denotes the bar level in projection, where it is necessary. This level is sometimes omitted. Several attributes in tags give syntactic and lexical information. Elements in bold blue will be displayed in the sentence viewer or can be used in other CALL applications. <PUNC> includes the punctuation.

If an error is found, a tag <ERROR> is marked into the first element <HEAD> which contains the error (e.g. agreement error). Then, if necessary, a tag <PARTERROR> is inserted into the second element which contains the error and is coindexed with the <ERROR> tag.

```
<LATLPARSE xml:lang="fr" option="freetext">
  <ANALYSIS complete="yes" errorDetected="yes">
    <PROJ cat="TP" tree="Ph">
      <PROJ cat="DP" attachedAs="spec" colorGramm="sujet" tree="GN">
        <HEAD cat="D" tool="yes" gender="masc" numbers="plu" lexeme="les"
          colorGrammLex="Det"> les </HEAD>
        <PROJ cat="NP" attachedAs="comp" tree="hidden">
          <HEAD cat="N" tool="no" gender="masc" numbers="plu" lexeme="chat"
            colorGrammLex="Nom">
              <ERROR index="Ab01" manypart="yes" category="NBR"> chats </ERROR>
            </HEAD>
          </PROJ>
        </PROJ>
      <BAR cat="T" colorGramm="pred" tree="GV">
        <HEAD cat="V" tool="no" number="sin" person="3" tense="present"
          mode="indicatif" lexeme="dormir" colorGrammLex="Verbe">
          <PARTERROR antecedent="Ab01"> dort </PARTERROR>
        </HEAD><PROJ cat="VP" attachedAs="comp" tree="hidden"/>
      </BAR> </PROJ> </ANALYSIS> </LATLPARSE>
```