

# Computerized linguistic resources of the research laboratory ATILF for lexical and textual analysis : Frantext, TLFi, and the software Stella

Pascale Bernard\*, Josette Lecomte\*, Jacques Dendien\*,  
Jean-Marie Pierrel \*

\* ATILF (Analyse et Traitement Informatique de la Langue Française, UMR 7118-CNRS/Université de Nancy2)  
44 Avenue de la Libération, BP 30687, F-54063-Nancy Cedex, France  
{pascale.bernard, josette.lecomte, jacques.dendien, jean-marie.pierrel}@inalf.fr

## Abstract

This paper presents some of the computerized linguistic resources of the Research Laboratory ATILF (Analyse et Traitement Informatique de la Langue Française) available via the Web, and will serve as a helping document for demonstrations planned within the framework of LREC 2002. The Research Laboratory ATILF is the new UMR (Unité Mixte de Recherche) created in association between the CNRS and the University of Nancy 2 since 2001-January 2nd, and succeeds to the local component of the INaLF situated in Nancy. This considerable amount of resources concerning French language consists in a set of more than 3400 literary works grouped together in Frantext, plus a number of dictionaries, lexis and other databases. These web available resources are operated and run through the potentialities and powerful capacities of a software called Stella, a search engine specially dedicated to textual databases and relying on a new theory of textual objects.

## 1. Introduction

Natural Language Processing is now one of the best answer to meet our societies' needs concerning Analysis and Extraction of Information.

Studies and Research projects in NLP more and more require large textual databases, either annotated or not, lexis and softwares. The cost of collecting these resources and making these tools is high enough to justify their pooling together and reusing to the mutual benefit of the community.

In this article, we present some of our textual resources and the software dedicated to them. They are accessible via the Web at <http://www.inalf.fr/atilf>

## 2. A set of linguistic resources for lexical and textual analysis

The two realizations we present below concern French language and are accessible via the Web. They are managed by a specially dedicated search engine called Stella which allows queries and hyper-navigation through and between the databases.

### 2.1. Frantext, a textual database

Frantext (1992) can be defined as a doublet constituted on the one hand by a vast corpus of written literary French texts mainly from the 16<sup>th</sup> to 20<sup>th</sup> centuries, and on the other hand by a software offering a Web interface with interrogation, consultation and hyper-navigation possibilities.

#### 2.1.1. History

Historically, the base was constituted in order to provide samples to be used in the elaboration of the TLF (Trésor de la Langue Française). It was started in the 60's. At that time, softwares were not interactive, and corpora were sequentially treated. In the 80's, a new approach prevailed : the creation of a kind of textual database platform which allowed direct access to the individual words of the corpus, and increased work efficiency. A first user's interface was realized in 1985, using telematic tools

such as Transpac and Minitel. About 90% of the 430000 samples cited in the TLF/TLFi are taken from Frantext.

Then, progressively, this first raison d'être turned to a new one : the desire to offer the scientific community a vast corpus of texts linked to a more and more efficient query tool.

#### 2.1.2. At the present time

At the present time, 3417 literary works are grouped together in Frantext. They cover a period dating from 1505 to 1998, plus one text dating back to 1377. There exists an other textual database covering the period 847 to 1502 (300 texts, soon accessible via the Web).

This corpus is subject to regular updating and enhancing, with three main objectives :

The first one is the good quality of the proposed texts. Some of them are still in the first state of their data capture (first 60's) and need to be amended.

The second one is the good quality of the editions considered for the data capture. Some of them can be considered as obsolete and/or not reliable.

And the third one is the enlargement of the database with new texts, in order to restore the balance between dates, or genres, or to facilitate some special operations in the domain of linguistic research or teaching.

#### 2.1.3. The two versions of Frantext

Frantext is accessible on the internet, on an annual subscription basis of 305€ The user will have access to the two versions of Frantext, which will be grouped together in a near future :

First, the total base, containing at present 3417 texts ( raw text, not annotated, for a total amount of more than 209 million occurrences). Texts can be interrogated on the graphic forms of the words (all texts at the same time, or one by one).

Second, a sub-base of the corpus, in "modern" spelling, containing 1940 texts (about 127 million occurrences), called "Frantext catégorisé". This corpus is morphologically annotated with Part-of-Speech labels, with a specific ATILF categorizer. Interrogation is possible on the graphic forms, and/or on the

morphological tags, either independently, or in the same request.

#### 2.1.4. The Frantext interface

The Frantext interface is user friendly and offers great possibilities :

Once the user has defined his work corpus (all the texts at the same time, or by author, or by dates, etc.), requests are possible. The user can search simple occurrences (words, tags ), co-occurrences, or sequences (possibly including optional terms), using simple graphic forms, word lists, or grammars. He can make a normal search, pointing on a word, a tag, a word-&-tag, a regular expression, or or make more advanced searches. He can save his results and have them frequency sorted, downloaded, etc.

There is a lot of help on line, explaining how to make a request, get results, or download a personal grammar, as well as modify the color of the screen or the window frame...

## 2.2. The TLFi, a lexical database

The computerized TLFi dictionary (Dendien, 1996) is the logical avatar of the TLF which was started in the 60's (with Frantext, its sub-project at that moment)

### 2.2.1. General overview

The TLFi (Trésor de la Langue Française informatisé) can be seen as a lexical database and a finely structured knowledge base. Its originality is based on its content : about 100000 words with more than 270000 definitions, special sections concerning their history, formation, etymology, more than 430000 examples and excerpts from the two last centuries literature.

### 2.2.2. Specificities of the TLFi

*The TLFi is specific in its content :*

Its word list is rich of about 100000 entries, all present in our funds and dictionaries. Also original, the treatment of morphemes (almost 60 words presented and defined under the headword -o), the treatment of prefixes, suffixes and other affixes.

It is structured regarding an over-elaborate list of meta-textual objects : headwords, grammatical codes, indications of domain, semantic and stylistic indicators, definitions, examples with their source... About 40 meta-textual objects.

Definitions are illustrated by a great quantity of examples : about 430000.

It proposes a great diversity of sections to be consulted: synchrony, etymology, history, pronunciation, bibliography, etc.

*The TLFi is specific in its structure :*

One of the main advantages of a computerized dictionary is to allow full-text requests throughout its whole content. However, in order to increase the precision and eliminate noise in the request, it would be useful to restrict a full text research to a specific kind of "textual object". That is the reason why the whole dictionary has been transformed into an XML document, with special delimiters for each type of textual objects.

A second dimension has been introduced : there is a hierarchy between the textual objects, using a special

internal tagset and a special control grammar. This makes possible to analyse the hierarchical structure of the article, and to take it into account when making a request.

### 2.2.3. Three levels of query

The TLFi can be freely consulted via internet, according three levels of query (depending on the user's needs).

A user can simply consult the dictionary, article after article, putting or not into evidence such or such type of information (a definition, an author, etc.).

He has a possibility to use a formulary of "aided request", i.e. consult the dictionary in a simple way (asking for a *definition*, a *domain* or another proposed "*object*"), or in a transverse way (crossing the criteria , for example : a *definition* in the *domain* of...).

A third way of consulting the TLFi is to use a more complex request crossing several criteria and taking into account the hierarchical structure of the textual objects. This request can be single- or multi-objects. It is possible to make and use word lists. For example, one can extract all the words ending with suffix *-âtre*, and then extract from this list all words having a *pejorative* meaning. One can also extract all the conjugated forms of the French verb *aimer* contained in the core of an example taken from Balzac, etc.

In conclusion, it has been proved that the fine structuring plus the rich content of the TLFi, allied to a very friendly user's interface allows very pertinent results when making requests.

## 3. Stella, a toolbox for the exploitation of textual resources

The two web available resources described above are operated and run through the potentialities and powerful capacities of a software called **Stella**, a search engine specially dedicated to textual databases and relying on a new theory of textual objects. A possibility of hyper-navigation exists between databases managed under Stella. Stella has been developed at the laboratory INaLF, now ATILF, by Jacques Dendien, to manage and exploit our textual resources.

### 3.1. Stella : a C++ toolbox offering several types of services

Stella offers developers three main types of service :

- *Web interfaces* for handling queries, user's sessions, dynamic menus, dynamic hyper-navigation between applications located either on the same server or not.
- *General services* such as data sorting, standard regular expressions, lexical databases with lemmatization and/or flexion of verbs, nouns and adjectives.
- *Management and exploitation of textual databases* : creation and maintenance of textual databases, optimal indexation system, open architecture (with abstract textual objects), high level query system.

Stella offers the users:

- A comfortable environment to make the requests. The interface is very friendly, with much help on line. It offers fine-grained

request possibilities, allowing precision in the results.

- An optimal response time to all requests.
- A good quality of service : Stella contains a linguistic “knowledge” (flexions, categorized databases) which allows a user to make complex requests.
- A powerful capacity of interrogation : a user can write parametrable grammars to be used and re-used in different contexts.
- A possibility of hyper-navigation throughout all the databases interconnected under Stella. For example, when consulting the TLFi, a user can “navigate” between Frantext, the TLFi and the French Academy dictionaries.

### 3.2. Examples of possible requests in Frantext

#### 3.2.1. The pronominal usages of a given verb : for example the verb “plaindre” (to complain) in Frantext

The main difficulties encountered concern the type of sequences it may be part of : affirmative constructions, negative or interrogative ones, simple or compound tenses. See below an example of a “grammar” which can help recognizing most of these pronominal usages in affirmative or negative sentences.

In the following grammar (we call it G1), comments are in italic, lines in bold correspond to “declarations of rules”. A rule XXX can be reused and called back into another rule using the syntax &rXXX. All rules must be declared, either above or below.

*[Rule describing discourse on the left of a pronominal verb in affirmative constructions.]*

**preambule\_affirmatif :**

je (me|m') | tu (te|t') | (se|s') | nous nous | vous vous

*[Rule describing a simple-tense affirmative construction (&cplaindre is for a conjugated form of the verb “plaindre”).]*

**temps\_simple\_affirmatif :**

&rpreambule\_affirmatif &cplaindre

*[Ditto for a compound-tense affirmative construction &cêtre is for an inflected form of the verb “ être ”]*

**temps\_compose\_affirmatif :**

&rpreambule\_affirmatif &cêtre &rparticipe\_passe

**participe\_passe :**

plaint | plainte | plaints | plaintes

*[Rule describing discourse on the left of a pronominal verb in negative constructions.]*

**preambule\_negatif :**

je ne (me|m') | tu ne (te|t') | ne (se|s') | nous ne nous | vous ne vous

*[Description of a simple-tense negative construction.]*

**temps\_simple\_negatif :**

&rpreambule\_negatif &cplaindre &rfin\_negation

*[Description of a compound-tense negative construction.]*

**temps\_compose\_negatif :**

&rpreambule\_negatif &cêtre &rfin\_negation  
&rparticipe\_passe

*[Ending terms of a French negation]*

**fin\_negation :**

pas|plus|jamais|guère|mie|point

*[All pronominal uses of a verb are described in one rule.]*

**usage\_pronominal:**

&rtemps\_simple\_affirmatif |  
&rtemps\_compose\_affirmatif | &rtemps\_simple\_negatif |  
&rtemps\_compose\_negatif

This grammar may be called in a request, by invoking one of its rules :

- &rtemps\_simple\_negatif calls the rule permitting the localization of pronominal uses of the verb in a negative construction, simple tense.
- &rusage\_pronominal calls the rule permitting the localization of all pronominal uses.

See below parts of the results obtained when calling this grammar on a Balzac sub-corpus of Frantext, thus showing the diversity of examples attested in Frantext:

#### a) Simple tense (affirmative form)

- **Je me plains**, qu'il n'y ait pas assez d'anecdotes, ne croyez pas que ce soit un vice de faiseur, mais goût de lecteur. (Balzac H. de / Correspondance T.3 / 1839)
- **Je me plains** de son abandon, elle m'appela fils dénaturé. (Balzac H. de / Le Lys dans la vallée / 1844)

#### b) Simple tense (negative form)

- Puis, quand tu ne m'aimeras plus, tu me laisseras, **je ne me plaindrai pas**, je ne dirai rien. (Balzac H. de / Histoire des Treize / 1835)
- Sachez –le bien, madame, je vous pardonne, et ce pardon est assez entier pour que **vous ne vous plaigniez point** d'être venue le chercher malgré vous...(Balzac H. de / Histoire des Treize / 1835)

#### c) Compound tense (affirmative form)

- Quand **je me suis plaint** de cette barbarie à un ami de M. Bellizard, il me répondit : Bah ! (Balzac Honoré de / Le Lys dans la vallée / 1844)
- Mais, Camille, je viens de reconnaître la vérité des critiques dont **vous vous êtes plainte** quelquefois. ( Balzac H. de / Beatrix / 1845)

#### d) Compound tense (negative form)

- ...Et **je ne me suis jamais plaint** ! (Balzac Honoré de / La Muse du département / 1843)
- ...j'ai murmuré, écrivait Pauline, mais **je ne me suis pas plainte**, Raphael ! (Balzac H.de / La Peau de chagrin / 1831)

This grammar can, of course, be completed for finding the verb *plaindre* in interrogative and interro-negative constructions. It is also possible to use parameters, allowing this grammar to be used for any verb.

#### 3.2.2. Enumerations and Repetitions (in POS-annotated Frantext)

Some writers use lists of items (adjectives on the left of a noun, adverbs, etc..). It is possible to detect and extract them with a Stella grammar. See below an example of parametrable grammar to be used to spot these various enumerations:

[The "item" rule defines the textual item which is going to be repeated. It contains two parameters : &1 et &2 which will be replaced by their real corresponding value when invoking the rule.

For example:

1) &ritem(en,S) invokes the item rule, passing it "en" et "S" as parameters. The item rule will then be equivalent to "en &e(g=S)" which means "en" followed by a noun.

2) &ritem(A) invokes the item rule passing it a first empty parameter and "A" as a second parameter. The item rule becomes equivalent to "&e(g=A)" which corresponds to an adjective.]

**item :**

&1 &e(g=&2)

[The "repetition" rule says that the textual item must be repeated twice (sub-expression "&ritem(&1,&2) , &ritem(&1,&2)" ) plus a number (greater than or equal to 1) of times (sub-expression "&+(, &ritem(&1,&2))" ). So, the item must be repeated 3 times or more.

This rule allows the user to search a textual item being repeated at least three times with an inserted comma between them.]

**repetition :**

&ritem(&1,&2), &ritem(&1,&2) &+(, &ritem(&1,&2))

In order to use this grammar (called for instance G2) the user will invoke the "repetition" rule passing it the two parameters which will thus be transmitted to the "item" rule. Here are some results obtained by this grammar on a corpus of some texts of Victor Hugo:

&rrepetition(A),G2 finds : " *sublime, simple, divers, profond, mystérieux, intime, fugitif*" (Les Feuilles d'automne); "*géographiques, politiques, moraux, intellectuels*" (Notre-Dame de Paris) ; "*ingrats, méchants, menteurs, jaloux*" (Les Rayons et les Ombres), etc.

&rrepetition(en,S),G2 finds : "*en musique, en mystère, en effroi*" (Les Quatre vents de l'esprit) ; "*en marbre, en granit, en jaspe, en porphyre, en velours, en satin, en pourpre, en drap*" (Le Rhin) ; "*en style, en art, en conscience, en idéal*" (Correspondance 1849-1866), etc.

&rrepetition(V),G2 finds : "*contemple, écoute, adore, aspire*" (Les Feuilles d'automne) ; "*flotte, ondule, bondit, tourbillonne*" (Notre-Dame de Paris) ; "*va, vient, rugit, hurle, mord*" (Les Contemplations), "*montait, descendait, lavait, brossait, frottait, balayait, courait, trimait, haletait, remuait.*" (Les Misérables) etc.

These two sample grammars intend to prove that all linguists, whether they are interested in syntax, semantics, or in stylistic researches can have a rather easy access to our resources, and find a real benefit in consulting them.

### 3.3. Examples of possible requests in the TLFi

#### 3.3.1. Aided request for getting all verbs, domain = religion, having a definition containing the graphic form faire :



The user has just to fill in the required slots, according to his need or curiosity. This request gives results such as the following:

Solution 1/15 : Article : ANNONCER, verbe trans. **verbe trans.**

**RELIG. CHRÉT.**

**Faire** connaître publiquement, prêcher comme un enseignement religieux.

Solution 2/15 : Article : CONFESSER, verbe trans. **verbe trans.**

**RELIG. CATH.**

C'est un aveu difficile à obtenir, une chose difficile à **faire**. (C'est le diable à confesser.)

Solution 3/15 : Article : DISPENSER, verbe trans. **verbe trans.**

**DR. CIVIL et RELIG.**

Autoriser (quelqu'un) à ne pas **faire** quelque chose de prescrit par une loi, une règle; accorder une dispense (cf. ce mot C). (Le maire peut dispenser des publications pour le mariage.)

Solution 4/15 : Article : DISPENSER, verbe trans. **verbe trans.**

**DR. CIVIL et RELIG.**

**Faire** remise à quelqu'un de ce qu'il a fait contre les règles de l'Église.

In order to get all definitions containing the verb *faire* as the first word of it, it is necessary to use complex requests (as in 3.3.2.).

#### 3.3.2. Complex request, for all entries ending with suffix -oir, the definition of which contains the word outil in position 1 :

First, it is necessary to use the facility *Gestion de liste de mots à partir des graphies du TLFi*:

Selection criterion = .\*oir

Name of the list = oir

This list contains 593 word forms. The user can read it, modify it if necessary before saving it and using it.

Then, the user fills in the formulary for a complex request, using this list. He asks first for a headword being

one of the list, and second, for a definition containing the given word “outil” at a distance of “+1” from the beginning of the definition (&d1)



He obtains 41 results, and see below the first four results of this request:

Solution1/41: Article : ACCORDOIR, subst. masc.

**ACCORDOIR**, subst. masc.

**Outil**, de forme variable suivant qu'il s'agit de traiter des cordes ou des tuyaux, servant à accorder les instruments de musique (piano, orgue, etc.).

Solution2/41: Article : AIGUISOIR, subst. masc.

**AIGUISOIR**, subst. masc.

**Outil** servant à aiguiser une lame ou tout instrument tranchant :

Solution3/41: Article : AMORÇOIR, subst. masc.

**AMORÇOIR**, subst. masc.

**Outil** utilisé par les artisans du bois, notamment les charpentiers, pour commencer les trous qui sont achevés ensuite avec des outils plus gros.

Solution4/41: Article : AVALOIR<sup>1</sup>, subst. masc.,

AVALOIRE, subst. fém.

**AVALOIR1**, subst. masc., AVALOIRE, subst. fém.

**Outil** servant à avaler la ficelle.

#### 4. Wide range of possible applications in various fields of research

It will be profitable to consult and navigate through the ATILF resources when research projects are related to :

##### 4.1. Collocations and Co-occurrences

It is possible to extract multi-word units, sequences of words, in order to determine which co-occurrences can be good collocations. It is not possible to constitute a concordance (neither from the TLFi nor from Frantext), but it is possible to extract lists of words present in a defined window, with their frequency. Though richly structured, the TLFi accepts full text research, just as any other data base.

A derived project about collocations and synonymy will start very soon.

##### 4.2. Sub-lexicon extraction

From Frantext, the user can select his own corpus according to various criteria (authors, dates, types of

texts), and create, for example the sub-lexicon of Victor Hugo, or of Colette.

In the TLFi, he can make requests according to other types of criteria: domains, grammatical tags (POS), and other types of textual objects: definitions, examples, etc. For example, he can extract lists of proverbs, lists of words of a specific domain, etc.

For example, there exists at ATILF a by-product of the TLFi: the lexicon *TLFnome*:

TLFnome96 : this lexicon, developed at InaLF by Marc Papin and Jacques Maucourt, contains 62.945 lemmas subsuming 389.524 graphic forms (1996 version; it is now being enhanced with about 27.000 new lemmas). To day, it is **not** distributed, but could be, if there is a demand for it.

Each entry (derived from the wordlist of the TLFi) is tagged with a particular tagset, this tagset being re-used for annotating the textual database Frantext.

#### 4.3. Morphological studies

By the way of word lists, a user can make requests about morphological phenomena, in Frantext as well as in the TLFi. Examples of derivation or composition phenomena, of multi-words containing a hyphen can be found in both resources. For example, in the TLFi, he can obtain the list of all transitive verbs ending with *-er* or *-oir*, or all adjectives ending with *-esque* or all verbs beginning with prefix *re-* or *dé-*.

#### 4.4. Local Syntax and Recurrent syntactic patterns

It is possible to study the environment of a word and have a better idea of the way it really functions:

For example the TLFi user can make a request about which type of object a verb accepts or not : Are there for example transitive verbs accepting a complementation including the preposition *de* ?

In Frantext, the contexts of a given verb can give information about its types of subjects, or objects, or modifiers.

In “Frantext catégorisé”, the user can get a list of *verbs* (tag = V) immediately followed by an “*infinitive verb*” (tag = Inf). For example, having selected from the base one text (Voyage au Congo, by André Gide) and thus built a restricted working corpus, this request applied on it gives 307 results. Among them: **allez chercher, pourrait dire, faudrait pouvoir, fait bouillir**, etc. The request can be more precise : If the first verb is restricted to occurrences of the verb *faire*, there are 92 results: **faire venir, faire fuir** etc. It is thus possible to have an immediate access to a list of attested examples of factitive or of modality constructions, etc.

#### 4.5. Semantic and Stylistic studies

If a user chooses to ask for the list of all adjective ending with *-esque*, present in the TLFi, it is perhaps because he wants to know which ones are *pejorative* or not...The request can be established, using this kind of criteria.

He can also make requests specific to an author of the many examples cited in the TLFi with their reference, and possibly ask which examples from Balzac contain a conjugated form of the verb *aimer*, not in the source of the

example, but in the core of the example. This is only possible because of the hierarchical structure of the textual objects in the article.

Frantext can be used by teachers and students in order to detect the nuances of a word, by looking at its environment. It is also possible to concentrate on the evolution of particular semantic fields, on the evolution of a word between its first attestation in the base and the last one, on problems linked to synonymy : are “synonyms” really “synonyms” (for example *suspicion* and *soupçon*, or *espoir* and *espérance*, etc.), and hence, what is “synonymy” ?

#### 4.6. Corpus tagging and Evaluation procedures

Part of the test corpus used in the GRACE French Part-of-Speech Tagging Evaluation Task (1994-1998) has been extracted from the Frantext database (by selecting texts without any copyright restrictions). The INaLF laboratory was a participant in this campaign (coordination committee, reflection committee) and had its Brill’s tagger for French evaluated at that occasion.

#### 4.7. Other fields of research

The TLFi contains special sections about Etymology, Pronunciation, and also numerous interesting remarks about Word Usage.

Frantext should be a prerequisite for all statistical studies about French words...

### 5. For information : Other freely accessible ATILF resources

The following resources are all accessible at our web site: <http://www.inalf.fr/atilf> (See : “tous les produits en accès libre” / Computer free access services):

- This year 2002 is very special: it celebrates the bicentenary of Victor Hugo’s birth. The laboratory ATILF propose a dedicated site by which you have access (among other products) to the specific **Hugo database** (derived from Frantext) containing 30 texts. One or two of them (“textes intégraux”) can be downloaded.
- The Historical Database for French vocabulary, called **BHVF** (“**Base Historique du Vocabulaire français**”). The 48 volumes of DDL (“Datings and Lexical Documents”) are computerized and can be queried, via Stella. The user can find in an only place a maximum of information concerning dating and “antedating” usually scattered about in ignored or uneasy to get documents. These data are taken from texts dating back to the 13<sup>th</sup> century. More than 61500 documents (among them 55000 “first attestations”) are proposed to the scientific community.
- **The WinBrill categorizer**, a Windows interface of Eric Brill’s tagger [Brill 1994] adapted for French at INaLF. The morphological tagging module has been put under Windows and can be freely downloaded at the ATILF site. The French parameters are freely distributed for research purposes, after signing a convention. **FLEMM** [Namer, 2000], is a lemmatizer which can be coupled to WinBrill : given a text

tagged with WinBrill, it affects each inflected form its lemma. Both tools can be downloaded.

- **Annotated Texts**: Also available, the on-line consultation of some texts which are no longer copyrighted (neither authors’ rights, nor editors’ rights). For example, Balzac’s *Cromwell*, *Bambara*, *Seraphita*, Beaumarchais’ *Le Barbier de Séville*, Laforgue’s *Les Complaintes*, Racine’s *Britannicus*. Some of them can be downloaded in XML format plus a Style sheet. (See also the ATILF-Victor Hugo site).
- “**Dictionnaires de l’Académie Française**”: The **French-Academy dictionary-8th edition (1932-35)** and the **French-Academy dictionary-9th edition (1<sup>st</sup> volume, 1992—from A to Mappemonde)** are freely accessible at our site, directly or by hypernavigation from Frantext or the TLFi .
- **Ancient Dictionaries**: Estienne (1552), Nicot (1606), Bayle (1740), French-Academy-1st edition (1694), French-Academy-5th edition (1798); French-Academy-6th edition (1835), etc.
- The laboratory ATILF also propose a site devoted to the **feminisation of French profession or trade name**.

### 6. Conclusion

Frantext, the TLFi as well as other computerized resources of the laboratory ATILF, do not compose a closed set of textual resources. They can be independently consulted and interrogated, and can be a starting point to a number of linguistic projects.

The objective of the laboratory ATILF is to let the community know that such resources exist for French Language. These resources have been initiated a long time ago, then developed at the laboratory INaLF, and are today available at ATILF, where work is still in progress.

In order not to exclude anyone of the process of distributing these tools, it seems interesting to propose a mutualization of these resources to the benefit of the entire community. The general policy of our laboratory is to welcome and give the research and teaching world the widest access to all our resources.

### 7. References

- Adda, G., J. Lecomte, J. Mariani, P. Paroubek, M. Rajman, 1997. Les procédures de mesure automatique de l’action GRACE pour l’évaluation des assignateurs de Parties du Discours pour le Français, in *Proc. of 1ères Journées Scientifiques et Techniques du Réseau Francophone de l’Ingenierie de la Langue de l’Aupelf-Uref, Avignon, Avril 1997*, pp. 245-257.
- Adda, G., J. Mariani, P. Paroubek, M. Rajman, J. Lecomte, 1999. Métriques et premiers résultats de l’évaluation GRACE des étiqueteurs morpho-syntaxiques pour le français, in *Proc. of la 6ème conférence sur le Traitement Automatique du Langage Naturel (TALN-99), Cargèse, (France), 12-17 Juillet 1999*, pp. 15-24
- Adda, G., J. Mariani, P. Paroubek, M. Rajman, J. Lecomte, 1999. L’action GRACE d’évaluation de

- l'assignation de parties du discours pour le français, *Langues*, 2(2), Juin 1999, pp. 119--129.
- Andreev, L., M. Olsen, 1999. Conception de systèmes hypermedia à grande échelle pour les sciences humaines ; Présentation de Philologic : le logiciel d'ARTFL, in *67th Congrès de l'ACFAS (Association canadienne-française pour l'avancement des sciences)*, May 11, 1999, University of Ottawa.
- Bernard, P., C. Bernet, J. Dendien, J.M. Pierrel, G. Souvay, Z. Tucsnak, 2001. Un serveur de ressources informatisées via le Web, in *Actes de TALN-2001, Tours, Juillet 2001*, pp. 333--338.
- Bonhomme, P., 2000. Codage et normalisation de ressources textuelles, in (*Pierrel, 2000*).
- Brill, E., 1993. *A corpus-based approach to Language Learning*. A dissertation in Department of Computer and Information Science, presented to the Faculties of the University of Pennsylvania.
- Brill, E., 1994. Some advances in Transformation-Based Part-of-Speech Tagging, in *Proceedings of the 12<sup>th</sup> National conference on Artificial Intelligence (AAAI-94)*.
- CNRS, 1976-1994. TLF, Dictionnaire de la langue du 19<sup>e</sup> et 20<sup>e</sup> siècle, Paris: CNRS, Gallimard.
- Dendien, J., 1991. Access to information in a textual database : access functions and optimal indexes, in *Research in Humanities Computing, Papers from the 1989 ACH-ALLC Conference*, Oxford: Clarendon Press.
- Dendien, J., 1996. Le projet d'informatisation du TLF, in *Lexicographie et Informatique*, pp. 25--34.
- FRANTEXT, 1992. Autour d'une base de données textuelles ; témoignages d'utilisateurs et voies nouvelles, Paris: Didier Erudition.
- Habert, B., A. Nazarenko, A. Salem, 1997. *Les linguistiques de corpus*, Paris: Armand Colin.
- Lecomte, J., 1995. Recommandations pour l'étiquetage morpho-syntaxique manuel de textes (document GRACE GTN-3-1.2, septembre 1995 [on the web : <http://limsi.fr/TLP/GRACE/pub1/doc>]).
- Lecomte, J., 1997, Codage MULTTEXT/GRACE pour l'Action GRACE-MULTITAG. Les étiquettes morpho-syntaxiques et les critères d'assignation (document GRACE-GTN-3-2.1, Décembre 1997 - [on the web : <http://limsi.fr/TLP/GRACE/pub1/doc>]).
- Lecomte, J., G. Adda, J. Mariani, P. Paroubek, M. Rajman, 1997. Progress Report on the GRACE Evaluation Program for French Part-Of-Speech Taggers, in *Proc. of the Speech And Language Technology (SALT) Club Workshop, Sheffield (UK), June 1997*, pp. 135--142.
- Martin, R., 2001. *Sémantique et Automate*, Ecritures électroniques, Paris: PUF.
- Namer, F., 2000. FLEMM : un analyseur flexionnel du français à base de règles, in *Traitement automatique des langues pour la recherche d'information*, CH. Jacquemin ed. TAL, 41(2)-2000, Paris: Hermès.
- Olsen, M., 1996. Text Theory and Coding Practice: Assessing the TEI, in *Joint Annual Conference of the Association for Computers and the Humanities and Association for Literary and Linguistic Computing, Bergen, Norway, June 1996*.
- Paroubek, P., J. Lecomte, G. Adda, J. Mariani, M. Rajman, 1998. The GRACE French Part-Of-Speech Tagging Evaluation Task, in *Proc. of the First International Conference on Language Resources and Evaluation (LREC'98), Granada (Spain), May 1998*, 1, pp. 433--441.
- Pierrel, J.M., 2000. *Ingénierie des Langues, Traité Information - Commande - Communication*, Paris: Editions Hermès.
- Rajman, M., J. Lecomte, P. Paroubek, 1997. Action GRACE, Format de description lexicale pour le français. Partie 2 : Description morpho-syntaxique, (document GTR-3-2.1, Juin 1997 -[on the web <http://limsi.fr/TLP/GRACE/pub1/doc>]).
- Véronis, J., 2000. Annotation automatique de corpus : panorama et état de la technique, in (*Pierrel, 2000*).