COMPUTATIONAL COMPARATIVE STUDIES ON ROMANCE LAGUAGES

A linguistic comparison of lexicon-grammars

Annibale ELIA & Yvette Mathieu

Istituto di Linguistica Università di Salerno
Laboratoire d'Automatique Documentaire et Linguistique
C.N.R.S. - Université de Paris 7

1. The linguistic comparison of lexicon-grammars

What we present here is an application on the basis of the italian and french linguistic data bank assembled by the Istituto di Linguistica of Salerno University (Italy) and the Laboratoire Automatique Documentaire et Linguistique (C.N.R.S.-France). These two researching centers have been working for years to the constitution of formalized grammars of the respective languages. The composition of lexicon-grammars is the first stage of this project.

A lexicon-grammar requires the study of nearly 300 syntactic property for all words of a language. The data revealed in this way are memorized under a matricial form of the type

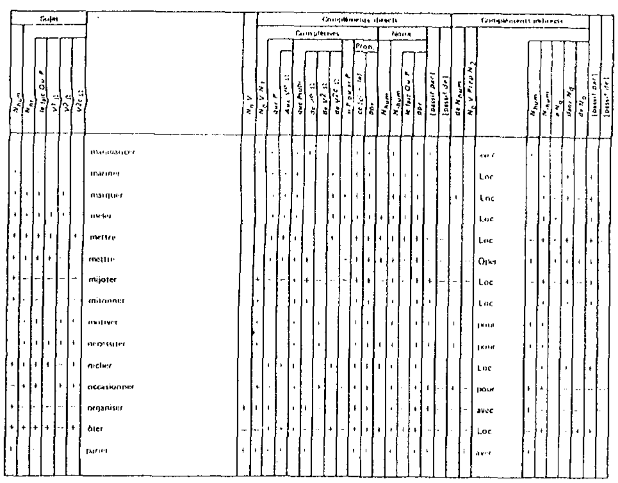

Fig.1

Therefore we have a number of classes of nouns, adjectives, adverbs, conjunctions and verbs, the determiners and the prepositions being included in the verbal classes.

Some matrices exist with lexicon-grammars correspondances between two lexicon-grammars as:

TABLE 45

N_0 =: Nhum	N_0 =: Nnr	N_0 =: il fatto Ch F	N_0 =: V1 Ω	Italien	N_0 è V-a	N_0 è V-pp Ω	di N_1 =: del fatto che F	di N_1 =: del fatto che Fcong	N_1 =: che F	N_1 =: che Fcong	di N_1 =: di V0 Ω	di N_1 =: di Aux V0 Ω	di N_1 =: se F o se F	di ciò	Ppv =: ne	di N_1 hum	di N_1 -hum	di N_1 contro N_2	sul fatto Ch F	a proposito del fatto Ch F	per il fatto Ch F	di N_1 per il fatto Ch F	con Nhum del fatto Ch F	a V0 Ω	Prep V0-inf Ω	V0-ndo Ω	Française	Table française
+	−	−	−	abbisognare	−	−	−	+	−	+	+	−	−	+	+	+	+	−	−	−	−	+	−	−	−	−	avoir besoin	−
+	−	−	−	abusare	−	−	+	−	+	−	−	−	−	+	+	+	+	−	−	−	−	+	+	−	+	−	abuser	8
+	−	−	−	approfittare	−	−	+	+	+	+	−	+	−	+	+	+	+	−	−	−	−	+	−	−	−	−	profiter	8
+	−	−	−	approfittarsi	−	−	+	+	+	+	−	−	−	+	+	+	+	+	−	−	+	+	+	−	−	−	en profiter	−
+	−	−	−	diffidare	+	−	+	+	−	+	+	+	−	+	+	+	+	−	−	−	−	+	−	−	+	−	se méfier	8
+	−	−	−	dimenticarsi	−	−	+	−	+	−	+	+	+	+	+	+	+	−	−	−	+	+	−	−	+	+	oublier	6
+	−	−	−	disperare	−	+	+	+	+	+	+	+	−	+	+	−	+	−	−	−	−	+	−	+	−	+	désespérer	8
+	−	−	−	disperarsi	−	−	+	+	+	+	−	+	−	+	+	+	+	−	−	+	+	+	+	+	+	+	se désespérer	8
+	−	−	−	dubitare	−	−	+	+	+	+	+	+	+	+	+	+	+	−	−	−	−	+	−	−	−	−	douter	8
+	−	−	−	esultare	+	−	+	+	+	+	−	+	−	+	+	+	+	+	−	+	+	+	+	+	+	+	exulter	8
+	−	−	−	fantasticare	−	−	+	+	+	+	+	+	+	+	+	+	+	−	+	−	−	+	+	−	+	+	rêver	8
+	−	−	−	favolare	−	−	+	+	+	−	+	+	−	+	+	+	+	−	+	+	−	+	+	−	+	+	rêver	8
+	−	−	−	favoleggiare	−	−	+	+	+	+	+	+	−	+	+	+	+	−	−	−	−	+	+	−	+	+	rêver	8
+	−	−	−	fidarsi	−	−	+	+	+	+	+	+	−	+	+	+	+	−	−	−	+	+	−	−	+	−	se fier	7
+	−	−	−	fottersene	−	−	+	+	+	+	+	+	+	+	+	+	+	−	−	−	+	+	−	−	−	−	se foutre	8
+	−	−	−	fregarsene	−	−	+	+	+	+	+	+	+	+	+	+	+	−	−	−	+	+	−	−	−	−	se ficher	8
+	−	−	−	gioire	−	−	+	+	+	+	+	+	−	+	+	+	+	−	−	+	+	+	+	+	+	+	jouir	8
+	−	−	−	godere	−	−	+	+	−	+	+	+	−	+	+	+	+	−	−	−	+	+	+	+	+	+	jouir	8
+	−	−	−	impazzire	−	+	+	+	−	+	−	+	−	+	+	+	+	−	−	−	+	−	−	−	−	−	raffoler	8
+	+	−	−	necessitare	−	−	−	+	−	+	−	+	−	+	+	+	+	−	−	−	−	+	−	−	−	−	nécessiter	10
+	−	−	−	pentirsi	−	−	+	+	+	+	+	+	−	+	+	+	+	−	−	+	+	+	−	−	+	+	se repentir	8
+	−	−	−	profittare	−	−	+	+	+	+	+	+	−	+	+	+	+	+	−	−	+	+	−	−	−	−	profiter	8
+	−	−	−	ricordarsi	−	−	+	−	+	−	+	+	+	+	+	+	+	−	−	−	+	+	−	−	+	+	se rappeler	6
+	−	−	−	ridersene	−	−	+	+	+	+	+	+	−	+	+	+	+	−	−	−	+	−	−	−	−	−	se rire	8
+	−	−	−	risentirsi	−	+	+	+	+	+	−	+	−	+	+	+	+	−	−	−	+	+	+	+	+	+	se ressentir	8
+	−	−	−	scordarsi	−	−	+	−	+	−	+	+	+	+	+	+	+	−	−	−	+	+	−	−	+	+	oublier	6
+	−	−	−	sognare	+	−	−	+	−	+	+	+	−	+	+	+	+	−	−	−	+	+	+	−	−	−	rêver	8
+	−	−	−	sovvenirsi	−	−	+	−	+	−	+	+	+	+	+	+	+	−	−	−	−	+	−	−	+	+	se rappeler	8
+	−	−	−	strafottersene	−	−	+	+	+	+	+	+	+	+	+	+	+	−	−	−	−	+	−	−	−	−	se contrefoutre	8
+	−	−	−	tacere	−	−	+	+	+	+	−	+	+	+	+	+	+	−	+	+	+	+	+	−	+	+	se taire	8
+	−	−	−	vergognarsi	−	−	+	+	+	+	+	+	−	+	+	+	+	−	−	+	+	+	+	+	+	+	avoir honte	−

Fig. 2

The composition of a lexicon-grammar allows to take into account selection and sub-categorization restrictions as well as derivational morphologic relations, without getting away from the limits of the formal syntax, that is, in limiting to the only experience of the estimation of acceptability, the recourse to intuitions of semantic nature.

A lexicon-grammar can satisfactorily render an account for the lexical and syntaxical differences of a word and those morphologically connected to it, allowing as well a comparison with verbs and nouns of another language disposing of a lexicon-grammar.

The high level of lexicon syntactic informations of the italian and french lexicon-grammars allowed us to a very subtle comparative study of the two languages. For exemple, the comparison of the italian verb **investire** and the french verb **investir**, leads to the following situation:

Italian

INVESTIRE 1

(1) NO V N1 di N2 =:

(1) a. La regina ha investito i Beatles del titolo di baronetti

(1) b. Il governo ha investito una commissione parlamentare del compito di indagare sul caso Moro.

INVESTIRE 2
(2) NO V N1 in N2 =:
(2) a. Max ha investito un milione nell'affare
(2) b. Eva ha investito (molto, tutta se stessa, il suo affetto) nel rapporto con Max

INVESTIRE 3
(3) NO V N1 (E,con N2) =:
(3) a. Il generale investì la piazzaforte (E, con le truppe d'assalto)
(3) b. Luca investì (Eva, l'auto di Eva) (E, con l'automobile)
(3) c. Eva investì Max (E, (con,di)(tutti i suoi problemi, parolacce))

INVESTIRE 4
(4) NO si V di N1 =:
(4) a. Lotario si investì del potere regale

INVESTIRE 5
(5) NO si V di N1 =:
(5) a. Max si è investito del (ruolo, personaggio) di Don Giovanni

INVESTIRE 6
(6) NO V N1 =:
(6) a. Questa difficile scelta investe tutti noi
(6) b. La trattazione investe anche i pronomi

INVESTIRE 7
(7) NO V Prep N1 =:
(7) a. La nave è investita (contro, su) lo scoglio

French

INVESTIR 1
(1) NO V N1 de N2 =:
(1) a. Le pape a investi l'évêque du titre cardinalice
(1) b. Le gouvernement a investi le ministre d'une charge importante

INVESTIR 2
(2) NO V N1 dans N2 =:
(2) a. Max a investi un million dans l'affaire
(2) b. Eva a investi (beaucoup, son affection) dans le rapport avec Max

INVESTIR 3
(3) NO V N1 (E, avec N2) =:
(3) a. Le général a investi la ville (E, avec ses troupes)

The french correspondents for the other uses of **investire** are:

INVESTIRE 3b. renverser, heurter, entrer en collision
INVESTIRE 3c. couvrir
INVESTIRE 4 s'approprier, usurper
INVESTIRE 5 entrer dans le rôle de
INVESTIRE 6a. toucher, concerner
INVESTIRE 6b. concerner
INVESTIRE 7 échouer

To give a comparative presentation of the italian and french data, we have two comparative tables of the realization of the morpheme INVEST- in italian and french. The first one (fig. 3) gives a survey of the correspondance, while a detailed comparison of the distributional syntactic properties and of the paraphrases is given by the second one (fig. 4)

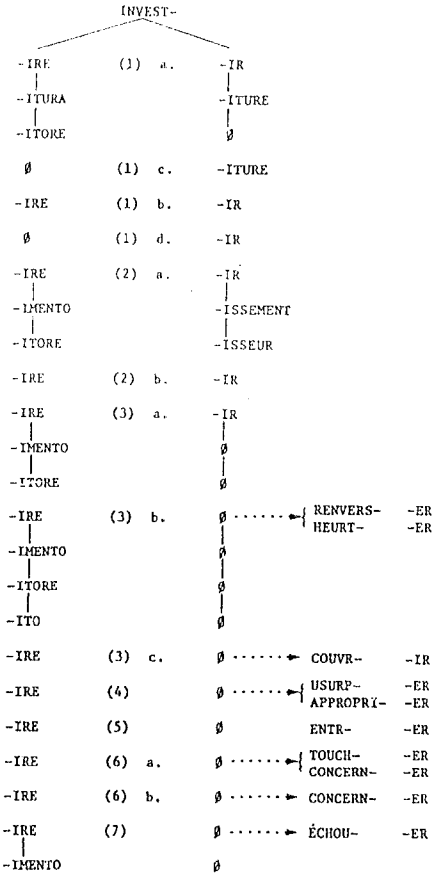

Fig. 3

INVESTIR fr.	INVESTIRE it.			DARE/DONNER -ITUR(A/E)	FARE/FAIRE -IMENTO/-ISSEMENT	PROCEDER/PROCEDER -IMENTO/-	ESSERE/ETRE -ITORE/-ISSEUR	ESSERE/ETRE -ITO/-I	
+	+	N_0 V N_1 (di/de) N_2 = : titre	(a)						
+	+	N_0 Vsup V-n (di/de) N_2 A N_1		+			+		
+	+	N_0 V N_1 (di/de) N_2 = : charge	(b)						
-	-	N_0 Vsup V - n (di/de) N_2 A N_1							
-	-	N_0 V N_1	(c)						1
+	-	N_0 Vsup V-n A N_1		+					
+	-	N_0 V N_1 (di/de) N_2 = : confiance	(d)						
-	-	N_0 Vsup V - n (di/de) N_2 A N_1							
+	+	N_0 V N_1 = : $ (in/dans) N_2	(a)						
+	+	N_0 Vsup V - n (di/de) N_1 (in/dans) N_2			+		+		2
+	+	N_0 V N_1 = : ψ (in/dans) N_2	(b)						
-	+	N_0 Vsup V-n (di/de) N_1 (in/dans) N_2			+				
+	+	N_0 = : militaire V N_1 (E+ ((con/avec) N_2))	(a)						
-	+	N_0 Vsup V-n (di/de) N_1 Comp				+			
-	+	N_0 V N_1 (E+ ((con/avec) N_2 = : voiture))	(b)						3
-	+	N_0 Vsup V - n (E+ ((con/avec) N_2))			+				
-	+	N_1 Vsup V - n					+	+	
-	+	N_0 V N_1 (E+ ((con/avec) N_2 = : métaph.))	(c)						
-	-	N_0 Vsup V- n (di/de) N_1 Comp							
-	+	N_0 si V di N_1							4
-	+	N_0 si V di Dét C_1 (di/de) N_2							5
-	+	C_1 = : *ruolo, personaggio* / rôle,...							
-	+	Nnr_0 V $Nhum_1$	(a)						
-	-	N_0 Vsup V-n (di/de) N_1							6
-	+	Nnr_0 V N-hum_1	(b)						
-	-	N_0 Vsup V-n (di/de) N_1							
-	+	N_0 V (E (Prép N_1))							7
-	+	N_0 Vsup V - n Comp			+				

Fig. 4

This table is to be read from the top left where the verbs **investire** and **investir** are placed. In the middle the properties are put in a line and noted by indications placed at the right concerning the number and the kind of use (example, (a)... 1. (b)..., 3, etc) At the top right, there are specifications on the support verb (Vsup) and the nominalization (V-n). The equivalent structures, italian-french, for prepositions and support verbs have a representation with a slash / (example di/de: **di**-italian and **de**-french; dare/donner: **dare**-italian and **donner**-french). At the left, there are two columns with plus and minus (+, -) which point out whether the property in question is valid or not for the verb; if the point is about the property with Vsup, first of all you must control if the verb at the left has a + or a - in the division of the column, if it is +, you can then read at the top right the type of Vsup and of V-n; if it is - , but there is a + at the right, that means that it refers only to one of the two verbs. The $ symbol is equivalent to "financial investments" and the symbol to "psychological investments".

Comparison between lexicon-grammars of different languages leads to draw up formal correspondences between structures, where so-called "grammatical words" are constants. Therefore, those "grammatical words" are considered as formal parts of structure. With that clause case, for example, we will have to take into account the following structures:

italian	french
...il fatto che F	le fait que P
... --	ce que P
...che F	que P

which will lead us to say that there is no italian correspondent to the french sentence

(1) **Max doute de ce que Marie vienne**

in view of the fact that the italian sentence

(2) ***Max dubita di ciõ che Maria venga**

is unacceptable. But the sentences

(3) **Max doute du fait que Marie vienne (fr)**

(4) **Max dubita del fatto che Maria venga (it)**

(5) **Max doute que Marie vienne (fr)**

(6) **Max dubita che Maria venga (it)**

drive us to consider italian structures (4) and (6) at a formal **proximity** degree of french structure (1). From a strictly formal point of view, the correspondence between (1) and (2) cant'be accepted, but from a larger point of view, which implies the formal proximity degree notion (obviously depending on the researcher's theoretical choice) we can mention structures (4) and (6) as the closest realization of the french structure (1). Of course, structures (4) and (6) have the strict formal correpondents (3) and (5).

The very complex data organization is usable by specialists but not by ordinary users. That is why we have been interested by the realization of a flexible and intelligent system, which allows the consultation and comparaison of the two languages lexicon-grammars (which are at the present day romance languages lexicon-grammars), and which includes the correspondance notion we have just presented. Its use will not need any knowledge neither of the data organization nor of computer science.

2. The TRANSLEGS system

The general TRANSLEGS's architecture is illustrated

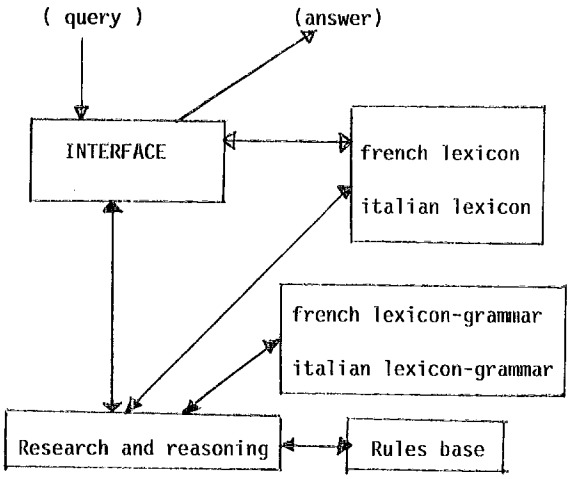

Fig. 5

A "query-answer" system of the interface allows to choose the type of research one wants to make. The system offers a research arborescence guided by its own resources and by the user's choices.

If the study concerns, for example, a class drawn up a know definitional property, TRANSLEGS will look for all the properties of this class, whether the correspondents exist in the other language, which elements appear (nouns, verbs, adjectives), in which structures do they appear and how do the correspondents of the other language behave.

If the study concerns a particular verb or noun, the system will look for the structures where this element appears, it will find its correspondent in the other language and the classes in which it exists, its properties and those of the correspondent into a given class. It will look for elements that have a degree of proximity for the secondary correspondence, and how do they behave in front of the properties and associated sentences.

In some cases, TRANSLEGS's answer needs deduction rules, if properties or classes exist in one language and not in the other one for example. Then, the system will propose "substitution" classes according to their formal proximity degree. The same type of answer will be done if one element has some properties in one language that its correspondent doesn't have.

The system being quite supple, the user can create his own environment

a) by modifying the deduction rules

b) by building new structures according to linguistic choices different from those reflected by existing classes (in that cas TRANSLEGS will make a class from the choosen definition property and those linked to it).

The system is written in PROLOG and runs on a VAX 11/750 from Digital Equipment.

3. Examples of dialogues

In the following examples, TL means TRANSLEGS

EX. 1

(TL) : Quelle est la langue de référence?

(User) : français

(TL) : Quel est le nom à étudier?

(User) : amour

(TL) : Voulez-vous
- la liste des classes où il apparait(1)
- étudier une classe particulière (2)
- étudier une classe particulière et connaître le comportement de l'équivalent italien (3)

(User) : 3

(TL) : Quelle classe voulez-vous étudier (numéro ou intitulé) ?

(User) : No Vsup Det N Prep N1

(TL) : Le nom **amour** possède les propriétés suivantes:
Verbe support: avoir

propriété N0 = Nhum équivalente à la propriété italienne N0 = Num

propriété Dét = **un modif** équivalente à la propriété italienne Det = **un modif**

propriété Dét = **du** équivalente à la propriété italienne Det = **del**

propriété Prep = **pour** équivalente à la propriété italienne Prep = **per**

Son équivalent italien **amore** possède les propriétés suivantes:
Verbe support: avere
...
...

Ex. 2

...

(TL) : Le verbe italien **derivare** possède les propriétés suivantes:

propriété 2 : soggetto non attivo

propriété 5 : completiva **dal fatto ch F**

Voulez-vous étudier son équivalent français (oui/non)

(User) : oui

(TL) : L'équivalent du verbe **derivare** est le verbe français **dériver**

Il possède les propriétés suivantes:

propriété 2: sujet non actif

propriété **3**: complétive **de ce Qu P**

La propriété 5 n'existe pas en français. La propriété 3 peut être considérée comme correspondante.

Ex. 3

...

(TL) : **ammirazione** avec **avere** a la propriété Dét = 0 **admiration** ne l'a pas

Suite proposée des propriétés correspondantes les plus proches:

Dét = **la**
Dét = **une modif**
Dét = **de la**

...
...

4. Conclusion

We would like Translegs to be computer tool for two main purposes:

- To allow an ordinary user to consult comfortably a linguistic data bank;
- To allow a linguistic researcher to work to a scientific description of a language or of two (or more) languages to compare by modifying and/or going deeper into lexical or grammatical informations already within the data bank made available by our laboratories.

REFERENCES

COLMERAUER, A. (1983) Prolog en 10 figures, Proc. of of IJCAI

DE ANGELIS, A.(1984) Nominalizzazioni con verbo supporto AVERE, Thesid Diss. unpubl.

ELIA, A., MARTINELLI, M., D'AGOSTINO, E. (1981) Lessico e strutture sintattiche, Napoli Liguori

ELIA, A. (1983) "Une hypothèse globale sur les verbes à compléments locatifs et sur l'infinitive locative", in Lingua e Stile 2, 1984

ELIA, A. (1984a) Le verbe italien, Bari- Paris Schena Nizet

ELIA, A. (1984b)"Sur l'unité du mot et la syntaxe comparée des langues romanes", in Revue Québécoise de Linguistique", 13, n.2

GROSS, M. (1975) Méthodes en syntaxe, Paris Hermann

GROSS, M. (1981) "Les bases empiriques du prédicat sémantique", in Langages, vol.63

LAMIROY, B. (1983)"La linguistique comparée et l'argumentation en syntaxe",in C.Angeltet et al. (eds.), Langue, dialecte, littérature, Leuven University Press

ROUSSEL, P. (1975) PROLOG: manuel de référence, Université Aix-Marseille 2

Vasseux, P. (1983) Système LEXSYN, L.A.D.L.,Université Paris 7