



HAL
open science

Combining a semantic map and a dual text corpus analysis for analysing verb polysemy (illustrated by French jouer)

Jacques François

► **To cite this version:**

Jacques François. Combining a semantic map and a dual text corpus analysis for analysing verb polysemy (illustrated by French jouer). Colloquium Ambiguity and disambiguation [GRAD 2006], May 2006, Sfax, Tunisia. pp.13-30. hal-04533998

HAL Id: hal-04533998

<https://hal.science/hal-04533998v1>

Submitted on 21 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

Jacques FRANÇOIS
Université de Caen & CNRS, France
jacques.francois@unicaen.fr

COMBINING A SEMANTIC MAP AND A DUAL TEXT CORPUS ANALYSIS FOR ANALYSING VERB POLYSEMY (illustrated by French *jouer*)

Verb polysemy may be tracked by listing the set of synonyms substitutable in a specific context to the item under examination. The *DES*¹ developed by the research lab CRISCO in Caen, which is based on about 400 000 synonymy links², delivers for each French source item a semantic map whose sense coordinates are “synonymy cliques”. The default semantic map is the best distributed bidimensional projection of a multidimensional semantic space with as many dimensions as synonyms of the source item.

For each synonymy clique and each predicate frame of the item, a rate of compatibility may be computed by evaluating the ability of each synonym to license each predicate frame of the source item in the same sense. On this basis, ‘semantactic’ areas i.e. areas exhibiting some syntactically as well as semantically close cliques – may be located on the semantic map of the source item. This method is to be illustrated by locating various predicate frames related to different senses of the highly polysemous French verb *jouer*³ such as

- Human subject**
 - ▶ *qn joue* NP<role, music work>
 - ▶ *qn joue* NP<game card, horse>
 - ▶ *qn joue de* NP<music instrument>
 - ▶ *qn joue à* INF (≈ pretends to do sth)
- Non animate subject**
 - ▶ *qch*<timber, hinge> *joue*
 - ▶ *qch*<argument> *joue*

A dual text corpus analysis based on the year 2002 of *Le Monde*, contrasted with recent French novels belonging to the text data-base FRANTEXT, suggests a rather similar profile of use for the various predicate frames in both corpora.

1. The semantic space of the first entry *jouer* (vs. *se jouer*) of the item JOUER as delivered by the *DES*

In the *DES*, the verbal item **JOUER** is split up into two entries *jouer* (intransitive as well as transitive) and *se jouer* (reflexive and middle). The first entry *jouer* is made up of 89 synonyms distributed among 7 components :

¹ *Dictionnaire Electronique des Synonymes*, under free consultation on the internet site www.crisco.unicaen.fr.

² These links are provided by 2 large dictionaries of the French language (Grand Larousse de la Langue Française) and Grand Robert), 2 synonymy dictionaries of the 19th century (by Guizot and Lafaye) and 3 synonymy dictionaries of the 20th century (by Bénac, Bailly and Bertaud du Chazaud). They were symmetrized, corrected and extended by Jean-Luc Manguin, manager of the *DES* at CRISCO.

³ For an extended syntactic and semantic study of the verb *jouer*, see Touratier (2000).

Component 1 : abuser, affecter, agir, agiter, amuser, attaquer, aventurer, badiner, batifoler, berner, blaguer, boursicoter, compromettre, contrefaire, copier, créer, donner, duper, enlever, exposer, exécuter, faire, feindre, feinter, figurer, flamber, flower, folâtrer, fonctionner, gratter, hasarder, imiter, incarner, influencer, interpréter, intervenir, manier, marcher, mentir, mettre en scène, mimer, miser, monter, mystifier, parier, passer, plaisanter, plastronner, ponter, poser, pratiquer, racler, railler, refaire, remuer, reproduire, représenter, rire, risquer, rouler, s'amuser, s'entraîner, s'exercer, s'ébattre, s'ébrouer, se divertir, se faire entendre, se mouvoir, secouer, simuler, singer, sonner, souffler, spéculer, taquiner, tenir, toucher, tourner, tripoter, tromper, user

Component 2 : avoir du jeu

Component 3 : coulisser⁴

Component 4 : gauchir, gondoler

Component 5 : jongler

Component 6 : pianoter

Component 7 : pincer

The 82 synonyms of the first component set up 90 cliques of synonyms. For instance, six cliques (1 to 6) include the synonym *abuser*. Out of these six cliques, the second and the third appear to be very akin, for both of them include 4 further synonyms : *berner*, *duper*, *mystifier* and *tromper*. Therefore, in the semantic map, these two cliques represent two very close sense coordinates in any bidimensional projection.

Clique 1	[2 syn.]	abuser, agir, JOUER
Clique 2	[6 syn.]	abuser , amuser, berner , duper , JOUER, mystifier , tromper
		\downarrow \downarrow \downarrow \downarrow \downarrow
Clique 3	[7 syn.]	abuser , berner , duper , JOUER, mystifier , refaire, rouler, tromper
Clique 4	[4syn.]	abuser, feindre, JOUER, mentir, tromper
Clique 5	[4 syn.]	abuser, JOUER, mentir, mystifier, tromper
Clique 6	[2syn.]	abuser, JOUER, s'amuser

The projection of the semantic space of the first component of the entry *jouer* looks like a reversed “Y” with three ends. At the superior end you find the synonymy clique {*jouer*, *miser*, *parier*, *ponter*} which is concerned with betting (for example on a horse in a horse race) : *Paul joue / mise / parie / ponte 50€ sur Walkyrie*. The left lower end is occupied by the clique {*contrefaire*, *copier*, *imiter*, *jouer*, *mimer*, *reproduire*, *simuler*, *singer*} which has to do with imitation, simulation or forgery. The object must convey a behaviour : *Marie joue / contrefait / simule / singe la surprise / la colère*, etc. Finally, the right lower end points out the clique {*badiner*, *blaguer*, *jouer*, *plaisanter*, *taquiner*} concerned with having fun usually with a partner introduced by *avec* : *Pierre joue / badine / blague / plaisante avec Suzanne*, or by a direct object : *Pierre taquine Marie* (\approx to tease) (see Figure 1).

⁴ For the exact computation of the interdistances between the synonymy cliques, see Ploux & Victorri (1998) and Victorri, Jacquet & Venant (2005).

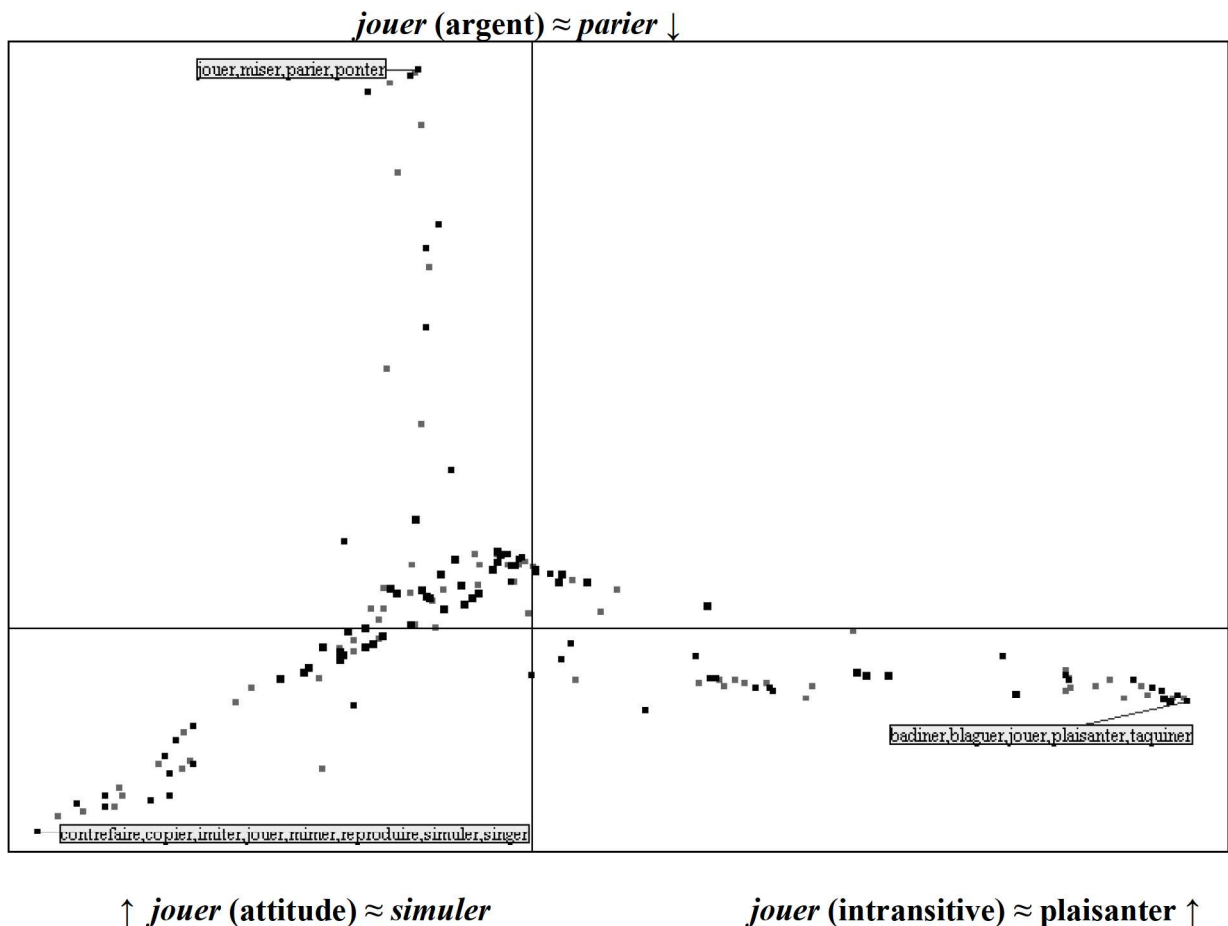


Figure 1 : The projection on the default semantic map (axes 1x2) of the entry *jouer* with its three furthest synonymy cliques

The number of close tied cliques at each end of the projection varies largely. At the upper end, only 2 cliques share at least 1 synonym with the extreme clique 83 and none of them shares more than 2 synonyms. The sense coordinate associated with clique 8 is therefore rather isolated :

83 {jouer,miser,parier,ponter}

▶ 2 synonyms shared :

84 jouer,miser,ponter,risquer

▶ 1 synonym shared

35 boursicoter,jouer,miser

At the right lower end, six cliques share up to 3 synonyms with the furthest clique 25:

25 {badiner,blaguer,jouer,plaisanter,taquiner}

▶ 3 synonyms shared :

24 badiner,blaguer,jouer,plaisanter,railler,rire

28 badiner,jouer,plaisanter,s'amuser,taquiner

▶ 2 synonyms shared :

26 badiner,folâtrer,jouer,plaisanter,s'amuser

27 badiner,jouer,plaisanter,railler,rire,s'amuser

33 blaguer,jouer,mentir,plaisanter

► 1 synonym shared :

29 batifoler,folâtrer,jouer,plaisanter,s'amuser

And at the left lower end, 10 cliques share up to 5 synonyms with the furthest clique 37 :

37 {contrefaire,copier,imiter,jouer,mimer,reproduire,simuler,singer}

► 5 synonyms shared :

38 contrefaire,faire,imiter,jouer,reproduire,simuler,singer

40 copier,imiter,jouer,mimer,reproduire,représenter,simuler

► 3 synonyms shared :

9 affecter,contrefaire,jouer,simuler,singer

39 contrefaire,feindre,imiter,jouer,simuler

56 faire,imiter,jouer,reproduire,représenter,simuler

► 2 synonyms shared :

8 affecter,contrefaire,feindre,jouer,simuler

55 faire,figurer,imiter,jouer,reproduire,représenter

► 1 synonym shared :

11 affecter,jouer,poser,simuler

59 faire,jouer,refaire,reproduire

61 figurer,incarner,jouer,reproduire,représenter

These quantitative data suggest a direct link between the number of synonyms of each furthest clique and the number of cliques sharing at least one synonym with it :

Furthest clique	synonyms	close cliques
83	3	2
25	4	6
37	7	10

Changing the projection plane generally introduces no more than a new furthest clique. Thus

- moving from plane 1x2 to plane 1x3 only displaces the clique 83 {*jouer, miser, parier, pointer*} from the upper to a lower middle zone, while introducing as the upper furthest clique Nr 20 {*amuser, berner, duper, flouer, mystifier, tromper*} typically referring to the domain of deception (see Figure 2),
- moving to plane 2x3 does not introduce any new furthest clique,
- and selecting plane 2x4 only replaces clique 20 by clique 49 {*enlever, gratter, jouer, râcler*}⁵ as a new upper furthest clique (see Figure 3).

⁵ This clique is obviously a "bridge clique", i.e. a heterogenous clique build up by mutually rather incompatible synonymy links : on the one hand *enlever, gratter* and *râcler* are synonyms of *jouer* as expressions of a bright (*enlever*) or poor (*gratter, râcler*) musical performance, on the other hand these three verbs are mutual synonyms in expressing (removing by) scraping, but then without link to *jouer*.

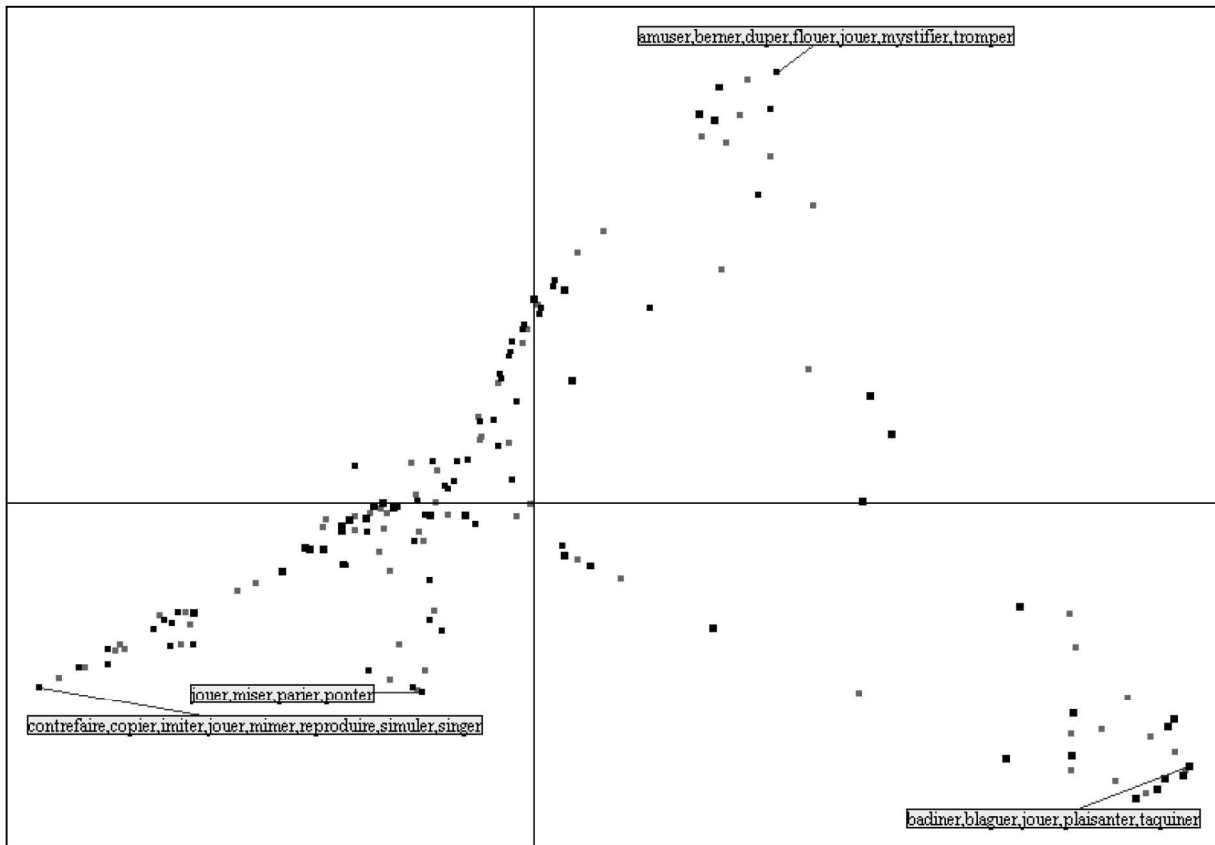


Figure 2 : The new upper furthest clique 20 on projection plane 1x3

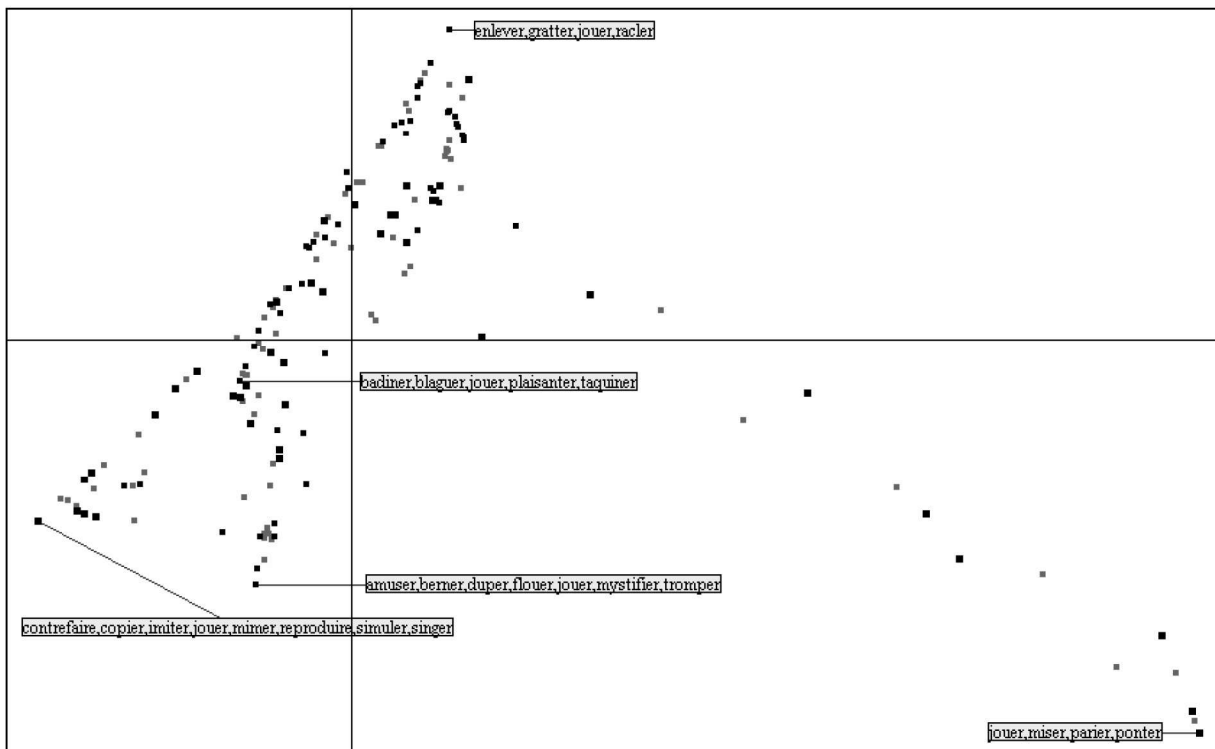


Figure 3 : The new upper furthest clique 49 on projection plane 2x4

2. Locating some semantactic areas in the semantic space of the entry *jouer*

For locating semantactic areas, the method applied here and in former papers (see François 2005, François to appear 1, 2) consists in

- i. listing all the predicate frames of the entry *jouer* ;
- ii. listing all the predicate frames of each of the synonyms of *jouer* ;
- iii. for a given predicate frame of *jouer*, listing all the verbs presenting a predicate frame agreeing semantically and syntactically with that of *jouer* ;
- iv. for each clique of synonyms, deciding on the basis of the occurring synonyms if *jouer* is meant in the meaning corresponding to the predicate frame in question ;
- v. for each clique of synonyms selected so, deciding how much synonyms fully (agreement rate : 1) or partially (agreement rate : 0,5) agree with the meaning and the predicate frame of *jouer* ;
- vi. locating on the semantic space of *jouer* the set – ideally a cluster – of cliques globally compatible with the predicate frame in question.

Table 1 displays 13 predicate frames associated to the verbe *jouer* and for each of its synonyms the list of compatible predicate frames. The most frequent frames under the synonyms of *jouer* are

<i>qn</i> ~	[s.o. is playing / having fun]
<i>qn</i> ~ <i>qch</i> (à un jeu / en bourse)	[s.o. is betting sth]
<i>qn</i> ~ <i>qn</i>	[s.o. is deceiving s.o.]
<i>qn</i> ~ à N ≈ feint d'être N	[s.o. pretends to be N]

Table 1 : The predicate frames compatible with each synonym of *jouer*

	qn joue	qch joue	qn joue qch <œuvre/partie>	qn joue qch (à un jeu/en bourse)	qn joue N<abs> (simule/figure)	qn joue qn = abuse	qn se joue de qn = se moque	qn joue à qch<jeu>	qn joue à N = feint d'être N	qn joue à INF	qn joue de qch <instr>	qn/qc joue un rôle	qn joue avec N<agent>
abuser	1					1							
affecter					1					1			
agir		1											
agiter											1		
amuser													
attaquer			1									1	
aventurer				1									
avoir du jeu													
badiner	1						1						
batifoler	1												
berner						1							
blaguer	1												
boursicoter				1									
compromettre						1							
contrefaire					1				1				
copier									1				
coulisser		1											
créer			1									1	
donner			1										
duper						1							
enlever			1										
exécuter			1										
exposer				1									
faire													
feindre					1				1	1			
feinter						1							
figurer												1	
flamber				1									
flouer						1							
folâtrer	1												
fonctionner		1											
gauchir		1											
gondoler		1											
gratter											1		
hasarder				1									
imiter					1				1				
incarner												1	
influer		1											
interpréter			1									1	
intervenir		1											
jongler													1
manier											1		
marcher		1											
mentir						1							

	qn joue	qch joue	qn joue qch <œuvre/partie>	qn joue qch (à un jeu/en bourse)	qn joue N<abs> (simule/figure)	qn joue qn = abuse	qn se joue de qn = se moque	qn joue à qch<jeu>	qn joue à N = feint d'être N	qn joue à INF	qn joue de qch <instr>	qn/qc joue un rôle	qn joue avec N<argent>
mettre en scène			1										
mimer					1				1				
miser				1									
monter													1
mystifier						1							
parier				1									
passer			1										
pianoter											1		
pincer											1		
plaisanter	1												
plastronner									1				
ponter				1									
poser									1				
pratiquer											1		
racler											1		
railler							1						
refaire													
remuer		1											
représenter			1									1	
reproduire					1				1				
rire							1						
risquer				1									
rouler						1							
s'amuser	1									1			
se divertir	1									1			
se faire entendre													
se mouvoir													
s'ébattre	1												
s'ébrouer	1												
secouer											1		
s'entraîner											1		
s'exercer											1		
simuler					1				1				
singer					1				1				
sonner			1										
souffler			1										
spéculer				1									1
taquiner							1						
tenir												1	
toucher											1		
tourner													
tripoter													1
tromper						1							
user													
Total	10	9	8	10	8	9	3	0	9	4	8	6	4

Three illustrations, abbreviated as A-C, are given below :

A. *qn joue* ≈ s.o. is playing / has fun

For 5 cliques of the first component, all constitutive cliques share the intransitive predicate frame with human subject *qn* ~ (agreement : 100%) and for further 3 cliques the agreement is 75%, meaning that on average 3 of 4 constituents of each clique are compatible with the frame (see Table 2). On the default projection (plane 1x2) the five cliques fully agreeing with the frame *qn* ~ are situated in the right lower area (the semantactic area concerned appears encircled on Figure 4).

Table 2 : The 8 cliques compatible with the predicate frame *qn* ~

	to banter	to frolic	to be joking	to frolic	to joke	to have fun	to amuse o.s.	to frolic	to frolic	Total	Number of Syn	Agreement rate of the clique with the predicate frame
	badiner	batifoler	blaguer	folâtrer	plaisanter	s'amuser	se divertir	s'ébattre	s'ébrouer			
<i>semantic agreement</i> →	1	1	1	1	1	1	1	1	1	9		
26 : badiner, folâtrer, jouer, plaisanter, s'amuser	1			1	1	1				4	4	100%
29 : batifoler, folâtrer, jouer, plaisanter, s'amuser		1		1	1	1				4	4	100%
30 : batifoler, folâtrer, jouer, s'amuser, s'ébattre		1		1		1		1		4	4	100%
63 : folâtrer, jouer, s'ébattre, s'ébrouer				1				1	1	3	3	100%
94 : jouer, s'amuser, s'ébattre, se divertir						1	1	1		3	3	100%
24 : badiner, blaguer, jouer, plaisanter, railler, rire	1		1		1					3	4	75%
25 : badiner, blaguer, jouer, plaisanter, taquiner	1		1		1					3	4	75%
28 : badiner, jouer, plaisanter, s'amuser, taquiner	1				1	1				3	4	75%

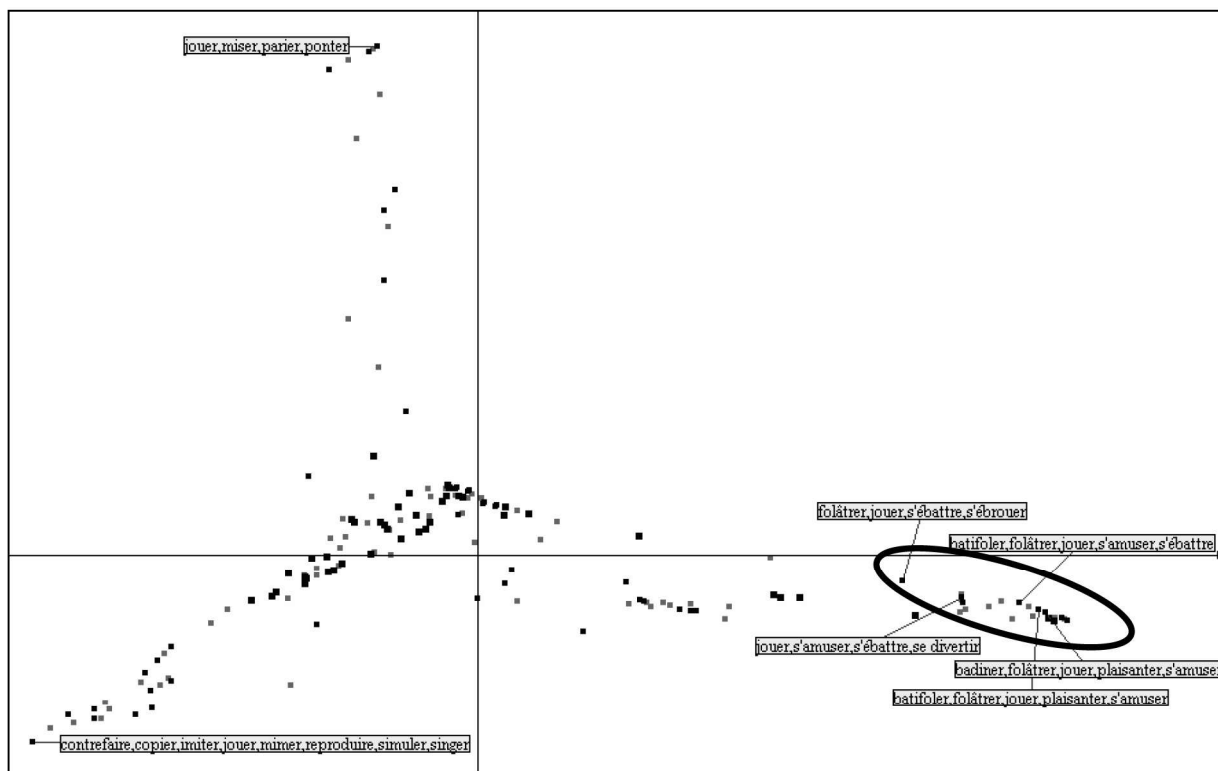


Figure 4 : Semantactic area of *qn joue* (s.o. is having fun) in the semantic space of the first component of the entry *jouer*

B. *qc joue* ≈ sth. works / warps

Three cliques of the first component fully agree with the intransitive predicate frame with inanimate subject *qc* ~. Furthermore, the components 2 {*avoir du jeu*}, 3 {*coulisser*} and 4 {*gondoler, gauchir*}, delivering respectively the three cliques 23, 41 and 68, are also associated with this frame (see Table 3). These six cliques appear closely assembled in the middle part on the semantic map of components 1-4 (Figure 5, default plane 1x2).

Table 3 : The 5 cliques compatible with the predicate frame *qch* ~

	agir	avoir du jeu	coulisser	fonctionner	gauchir	gondoler	influencer	intervenir	marcher	remuer	Total	Number of Syn	Total / Number of Syn
<i>semantic agreement</i> →	1		1	1	1	1	1	1	1	1	9		
15 : agir, fonctionner, jouer	1			1							2	2	100%
16 : agir, influencer, jouer	1						1				2	2	
17 : agir, intervenir, jouer	1							1			2	2	
23 : avoir du jeu, jouer		1									1	1	100%
41 : coulisser, jouer (Comp.3)			1								1	1	
68 : gauchir, gondoler, jouer (Comp.4)					1	1					2	2	

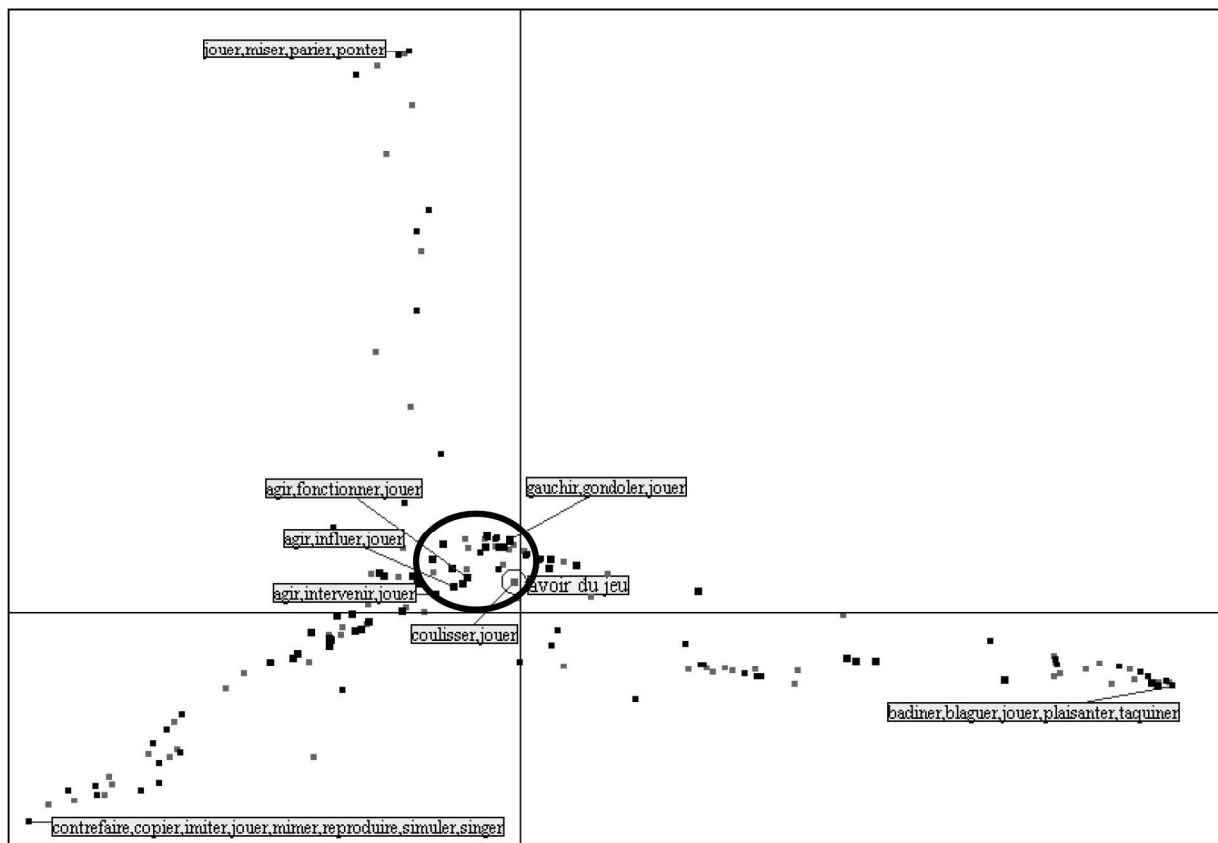


Figure 5 : Semantactic area of *qqc joue* in the semantic space of the components 1-4 of the entry *jouer*, i.e. including : *coulisser* (to slide), *avoir du jeu* (to be loose), *gauchir / gondoler* (to warp)

C. *qn joue qc*<work> ≈ s.o. is playing / performing sth.

Here, the average agreement level between synonyms and predicate frame is lower. Therefore cliques with an inferior agreement of 67% are taken into account. The reason for this low agreement is that only 4 synonyms {*donner, exécuter, interpréter, représenter*} are associated with this frame in the domain of musical entertainment and show business. Semantically, the other seven synonyms only partially agree with the frame (see Table 4). In Figure 6 (default plane 1x2), the eight concerned cliques appear quite close assembled in the middle of the left lower branch.

Table 4 : The seven cliques partially agreeing with the frame *qn joue qc* <œuvre/partie>

	to launch into	create / produce	perform	perform con brio	perform	perform	to stage / direct	to show	to perform / play	to sound	? to sound	Total	Number of Syn	Total / Number of Syn
	attaquer	créer	donner	enlever	exécuter	interpréter	mettre en scène	passer	représenter	sonner	souffler			
<i>semantic agreement</i> →	0,5	0,5	1	0,5	1	1	0,5	0,5	1	0,5	0,5	7,5		
47 : donner, jouer, passer			1					0,5				1,5	2	75%
82 : jouer, mettre en scène, représenter							0,5		1			1,5	2	
44 : donner, exposer, jouer, représenter			1						1			2	3	67%
45 : donner, faire, jouer, représenter			1						1			2	3	
53 : exécuter, faire, interpréter, jouer					1	1						2	3	
57 : faire, interpréter, jouer, représenter						1			1			2	3	
71 : incarner, interpréter, jouer, représenter						1			1			2	3	
42 : créer, faire, interpréter, jouer		0,5				1						1,5	3	

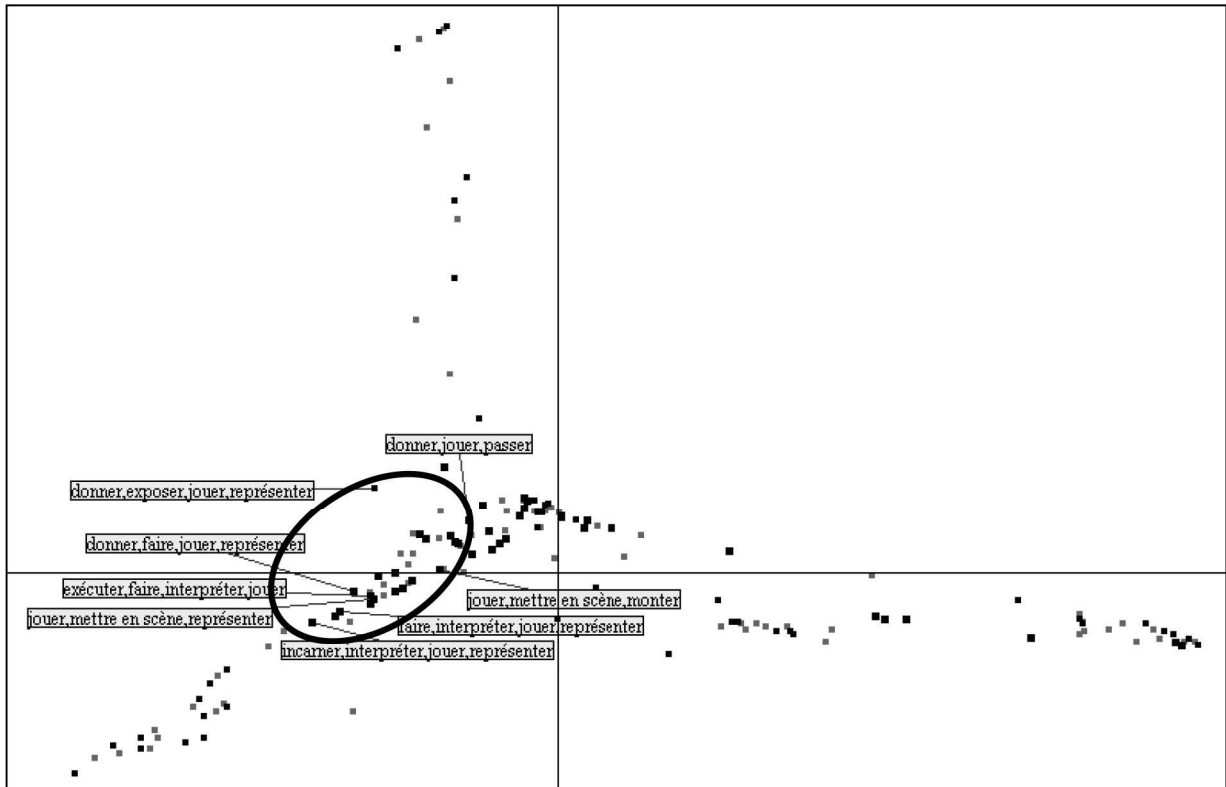


Figure 6 : Semantactic area of *qqn joue qc* <work> in the semantic space of the 1st component of the entry *jouer*

2. Contrastive text corpus analysis

Finally I would like to search the place occupied in different text types by the three predicate frames studied. The comparative method adopted consists in randomly establishing two lists of 100 occurrences of the verb *jouer*. The first list is made up by occurrences coming from the year 2002 of the daily paper *Le Monde* (polished journalistic style) and the second by occurrences borrowed from the most recently entered novels (of the nineties) in the categorized text database FRANTEXT⁶ (see Table 5).

The A frame (*qn joue* ≈ s.o. is playing / has fun) holds the first rank in the daily paper corpus with 30 occurrences, e.g. :

- (1) Je venais jouer là avec mes camarades , on se cachait derrière les stèles . [30-12-2002]

but only the third rank in the novel corpus with 14 occurrences, e.g. :

- (2) En revenant de l'école, Nicolas faisait ses devoirs ou jouait avec son petit frère en prenant soin de ne pas faire de bruit. [E. Carrère, *La classe de neige*, 1995, p. 37]

But this gap of 16 occurrences is compensated for by the opposite gap concerning the frame *qn joue à qc* <jeu> (s.o. is playing a game) which holds the first rank in the novel corpus with 24 occurrences against only 7 in the daily corpus (5th rank). Thus, on the whole, the play activity proves to hold the first rank in both text types, but in the novel corpus the type of game played is more frequently specified.

The two other frames studied collect a rather similar number of occurrences in both text types. In the daily corpus, the B frame (*qc joue* ≈ sth. works / warps) holds the seventh rank with 6 occurrences, e.g. :

- (3) Le temps joue en notre faveur. [31-12-2002]

and the fifth rank in the novel corpus with 7 occurrences, e.g.

- (4) Ce maudit blocus joue contre nous, je vous le dis ! [P. Rambaud, *La bataille*, 1997, p. 22]

Finally, the C frame (*qn joue qc* <work> ≈ s.o. is playing / performing sth) holds the third rank in the daily corpus, e.g. :

- (5) L'Eglise catholique compte une diversité de réalités et d'organisations qui jouent chacune leur partition . [26-12-2002]

and the second rank in the novel corpus with 24 occurrences, e.g. :

- (6) Là aussi se jouait une drôle de farce... [A Boudard, *Mourir d'enfance*, 1995, p. 101]

⁶ Accessible by suscription on the site of the lab ATILF, CNRS in Nancy

	<i>LM 2002</i>		<i>FRANTEXT</i>	
	Freq	Rank	Freq	Rank
<i>qn joue</i>	30	1°	14	3°
<i>qn/qc joue Det rôle</i>	21	2°	14	3°
<i>qn joue qc <work></i>	17	3°	24	2°
<i>qn joue qc (≈ to bet ; gamble)</i>	9	4°	1	9°
<i>qn joue à qc<jeu></i>	7	5°	25	1°
<i>qn joue de qc <instr></i>	7	5°	5	7°
<i>qc joue</i>	6	7°	7	5°
<i>qn joue N (≈ to affect ; feign)</i>	2	8°	7	5°
<i>qn joue à INF</i>	1	9°	2	8°
TOTAL	100		100	

Table 5 : Frequency and rank of 9 predicate frames of the verb *jouer* in a daily and a novel corpus

In conclusion, this brief study shows that the *DES* is able to deliver bidimensional projections of the multidimensional semantic space of lexical units – and here especially of French polysemous verbs – on which specific predicate frames prove to occupy a more or less restricted area. I called such areas “*semantactic*” because the cliques concerned share syntactic as well as semantic properties. One is tempted to assume that in different text types, the nine types of predicate frames distinguished may hold a various frequency rank. But in the first tentative comparative study proposed here between a daily and a novel corpus of a similar period, no striking gap emerged. Dividing the daily corpus into domains (politics, economy, sport, etc.) might reveal frequency differences which are imperceptible in the whole daily corpus.

References

- François, Jacques (2005,) « Polysémie verbale et cadres prédicatifs : *demander* et ses synonymes ». in F. Lambert & H. Nølke (dir.), *La syntaxe au cœur de la grammaire*. Collection *Rivages linguistiques*. Presses Universitaires de Rennes, p.85-98
- François, Jacques (to appear 1), *Pour une cartographie de la polysémie verbale*. Leuven : Peeters, Bibliothèque de la Société de Linguistique de Paris.
- François, Jacques (to appear 2), « Quand jouer, c’est jouer de la musique : Repérage contextuel de quatre zones de l’espace sémantique du verbe ‘jouer’ ». In E. Richard & F. Neveu (dir.), *Actes du colloque de Rennes en hommage à Michèle Noailly, octobre 2005*. Not specified publisher.
- Ploux, Sabine & Bernard Victorri (1998), « Construction d’espaces sémantiques à l’aide de dictionnaires de synonymie », *TAL* 39, p. 161-182.
- Touratier, Christian (2000), *La sémantique*. Paris : Armand Colin, Collection Cursus.
- Victorri, Bernard, Guillaume Jacquet & Fabienne Venant (2005), « Polysémie lexicale », in : P. Enjalbert (ed. 2005), *Sémantique et traitement automatique du langage*. Paris : Hermès-Lavoisier, p.99-132