



**HAL**  
open science

# ANALYSING JOURNALISTIC DISCOURSE AND FINDING OPINIONS SEMI-AUTOMATICALLY?

Fabienne H. Baider

► **To cite this version:**

Fabienne H. Baider. ANALYSING JOURNALISTIC DISCOURSE AND FINDING OPINIONS SEMI-AUTOMATICALLY?. 2014. hal-00919370v2

**HAL Id: hal-00919370**

**<https://hal.science/hal-00919370v2>**

Preprint submitted on 16 Apr 2014 (v2), last revised 2 May 2014 (v3)

**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.

**THIS PAPER HAS BEEN SENT FOR PROFESSIONAL EDITING.**

**THIS IS ONLY A DRAFT.**

ANALYSING JOURNALISTIC DISCOURSE AND FINDING OPINIONS SEMI-AUTOMATICALLY?  
A CASE STUDY OF THE 2007 AND 2012 PRESIDENTIAL FRENCH CAMPAIGNS

Baider Fabienne  
University of Cyprus  
[fabienne@ucy.ac.cy](mailto:fabienne@ucy.ac.cy)

### **Abstract**

The goal of this paper is to test three different NLP technologies to analyze French journalistic discourse about candidates during French presidential campaigns in order to evaluate discourses differences depending on the candidate gender and/ or the candidate political party. Indeed it is suggested that working on a journalistic corpus with specific software can help studying linguistic patterns and choices which are made on the basis of political affiliation or gender stereotypes. These conclusions are drawn from quantitative and qualitative analysis carried out with 1. the software SEMY which gives semantic profiles semi-automatically; 2. the software ANTCONC which provides useful *Keyword in Context (KWIC)* or abstracts of the text in which is used the studied item, as well as collocations; 3. the software TERMOSTAT which works on discourse specificities, frequencies and most used morpho-syntactic patterns. Convergent asymmetries between female and male candidates in journalistic discourse (however conditionally) were found as far as our data dedicated to the 2007 and the 2012 presidential campaigns are concerned. We conclude that social gender (i.e. stereotypical expectations about who will be a typical member of a given category) and / or political favoritism may affect the representation of leadership in discourse and may affect in turn the readership, hence the electorate. Thus the paper recommends the use of corpus linguistic tools to support semi-automatic investigation of characteristics of political texts.

### **Introduction**

Work analyzing ideology in discourse (Baider & Constantinou 2013, Kohl 2011, Reutenauer *et al.* 2010; Valette 2010; Mussolff 2008) showed the indexicality (indirectness) of the construction of prejudices in discourse, i.e. *the lexical and stylistic choices* can give an ideological (racist or sexist) semantic prosody to a text without the text being explicitly racist or sexist. Therefore *lexical specificities* characterizing a discourse can be a mean to detect ideological position. The purpose of this article is to show how indeed software such as ANTCONC, SEMY or TERMOSTAT (Drouin 2003, Lesage *et al.* 1993) can help in finding the ideological positioning of discourses. On the basis on different studies based on journalistic data (2007 and 2012 election campaigns in France), we show that results found in a semi-automatic way allow a fairly good comparison of political profiles and campaigns. After a brief review of research within the interface, language, media and gender issues we present our data and tools in our second section. Our data (journalistic discourse) are focused

on specific candidates participating in the presidential campaign; the tools we use can identify lexical preferences (NP extractor, concordance and syntactic parser). The last section analyses the results and argues that such software help to verify, invalidate or refine results obtained with previous qualitative studies or manual readings: they offer complementary results to detect lexical and stylistic specific features. We do however warn that the results and the interpretations have to be confronted with a close manual reading and a good knowledge of the social context, so that we can verify whether the findings are plausible and how to decide between different interpretations of the lexical findings.

## 1. Previous works

According to van Dijk's theory (2006 *et passim*) the overall strategy of ideology consists in a dichotomous positioning between several entities:

- positive presentation /action of our preferred candidate, of the party, our electorate, etc., emphasizing their good things, de-emphasizing their bad things;
- negative / action of the other champion candidate, of the other party or of the other electorate, etc., emphasizing their bad things and de-emphasizing their good things.

In order to position oneself along a certain ideological point of view, different linguistic devices are used such as choice of lexical units, grammatical constructions, metaphors, irony, etc. Indeed this dichotomous positioning can be done in the most indirect fashion.

### 1.1 Indexicality of Ideology and linguistic choices

As mentioned in our introduction, ideology in discourse is mainly indexical (Van Dijk 2006, Salim 2011). Therefore Vallette & Rastier (2006, 4) identified racist content in online documents by observing language choices which had for results the writing of a text with the illocutionary force of a racist text without being racist explicitly. For instance as far as morpho-syntax is concerned, indexical racism was characterized by an intensive use of evaluative adverbs and negative adverbs (*nothing, never, much, too much*, etc.) and an intensive use of first person plural pronoun in order to build an opposition between <us> and the <other(s)>, as well as a predominant usage of pejorative morphemes, among the observed linguistic patterns.

In the same line of thought, previous research on sexism in discourse noted that language choices can indeed build semantic oppositions and discourse asymmetry between the described entities, without doing so in an obvious manner (Ochs 1992, Kitzinger 2007, Goodwin 2011). Working on sensationalist discourse (such as in the *Sun*), Clark (1998) observed the shifting of blame in sexual violence onto the victims by analyzing 'naming' and 'transitivity'<sup>1</sup> (1998, 183-185). She concluded that different connotations of approval and / or legitimacy are carried out by these linguistic patterns. In education, previous studies (Lazar 2005, Penelope 1990) have shown that in many English-language textbooks most of the agents or doers are men and boys or that examples chosen depict unfavorable traits for female human beings. Bergvall (1996) and Macaulay & Brice (1997) have concluded that the examples invented by linguists to illustrate the phenomena of syntax were characterized by an

---

<sup>1</sup> Transitivity is analysed to decode lack of responsibility or blame, this is carried out by taking into account the process, the participants in the process and the circumstances of the process (see the framework of Halliday, 1985).

under-representation for female beings especially in examples indicating agency (9%) as well as a negative semantic prosody since they presenting women as passive, incompetent and talkative. This was despite the recommendations from the LSA (Linguistic Society of America) regarding the construction of examples.

Similarly Michard (1988) and Michard and Ribery (2008), working from scientific texts in French, concluded that female human beings were constructed in that discourse as 'non-agent' entities (2008, 174-184), because of the choice of verbs or nouns referring to them; they were also evaluated with another entity (high number of adverbs and adjectives such as *few*, *a little*, *a few*, *less* collocated with the Noun Phrases under investigation). Combining their syntactic and lexical findings, Michard and Ribery concluded that a power relationship can be established discursively between the described entities.

Indeed such an asymmetry may explain the functions and the qualities assigned to one or the other person within concrete social situations such as the political scene.

### 1.2 *Journalistic discourse, leadership and indexing gender*

Many studies on leadership have concluded that gender play an important role in the perception of a public *persona*. Some conclude that even though the negative perception of female politicians by their peers can be attributed to a complex interplay of factors, it can also be partly attributed to the way female politicians draw (if they do) on gendered linguistic resources (Shaw 2012, 39); others mention that similar verbal behaviours for male and female managers will be viewed rather negatively for women (Sung 2013; Thimm et al. 2003).

As far as media are concerned, Cotter (2011) studied two key parameters in news media in relation to changes in women's role and visibility in society: bylines (who is entitled to be recognized by name when writing an article) and story topic (what counts as salient in the journalist's world). Both are elements of prestige in the journalistic discourse community. Her conclusions is that 'despite increased opportunities for women over time, the position occupied by women relative to men, family, and the economy has not changed to the degree we sometimes think it has'. This in turn can partially explain the gender disparities observed in journalistic discourse.

As a matter of fact a discursive asymmetry was noted relating to female and male politicians (Kahn 1994, 1996; Sreberny-Mohammadi & Rossi 1996; Roncarolo 2000; Coulomb - Gully 2009 and 2012; Bertini 2007). Most studies focused on French politicians in written and oral corpus raised the issue of the role of gender, the gender of the entity described as well as the gender of the entity writing the text. Thus Caillat (2012), Rabatel (2012) and Devriendt (2012) showed how the reported speech in journalistic discourse was adapted differently when comments were referring to a female candidate or to a male candidate in the 2012 presidential campaign. Similarly, Barnes & Larrivé (2011), Moores (1997) Baider (2009) examined how female and male politicians are constructed in the French newspapers. They have concluded that discourse practices and lexical choices play a critical role in the construction of the entities described as incompetent or in the opposite as being the perfect leader.

However the ideological distance between the newspaper and the candidate suggests that power struggles are a primary source of asymmetrical treatments as much as gender (Barnes & Larrivé 2011; Baider 2009, Moore 1997): Barnes & Larrivé

(2011, 2507) concluded that indeed a right-wing newspaper refers to the left-wing female politician (Arlette Laguiller) by the familiar forename alone because of her political affiliation and not because of her gender. Such dis-favorable treatments are though dis-preferred when no competition is at stake (Moores 1997, Baider 2009).

The general conclusion drawn in these previous studies is the indexicality of ideological positioning, this indirectness renders (semi)-automatic reading difficult. Indeed how could we evaluate such positioning in discourse? Since the studies listed and described above point at *lexical specificities* characterizing a discourse taking an ideological position, we could then focus our analysis on the most frequent and specific lexical units in our data to test the hypothesis that these units will characterize a certain point of view.

### 1.3 *Indexicality and corpus linguistics*

Much research on political analysis have been using software to extract lexical units in order to find some patterns, which could then in turn explain either the motivation of the word choice, either the impact of the word choice. They all conclude that rhetoric and word choices are fundamental in the support of political strategies. For instance political journalists accredit this view by showing political life primarily through quotations or small excerpts, leaving aside much of the rest of the speech (Bart 2010; Saita 2008). Brasart (1998), in his study of ‘petites phrases’ in French political discourse, demonstrated the existence of linguistic routines which politicians select in order to make an impression on the audience. Semi-automatic readings allow us to examine how and to what degree politicians fit into their strategy rhetoric such expressions; they know that their words will often be reduced to a few sentences and they maybe anticipate and premeditate the reduction journalists will be doing (Krieg Planke 2011).

Quantitative observation also helps to evaluate changes in political speeches, avoiding the axiological dimension which frequently marks both the discourse and its analysis. Corpus analysis using semi-automatic reading can also assess the evolution of the professionalization of communicative labor policies and underlying standards expressible in politics (Marchand, 1998, Sinclair 2004). For instance Marchand (2009) showed that computer-assisted discourse analysis (statistical text analysis, morphosyntactic analysis) can contribute to the diagnosis of genre, style and attitude in language. The study of the French prime ministers statements throughout the Fifth Republic (Marchand & Dupuy 2013) showed that the frequency and specificity criteria can identify the shift in lexical choices; the study concluded that this change in word choices would also reflect a change in priorities for the presidential debates in France: moving away from speeches centered on diverse topics all close to the social fabric such as education and environment, French debates since the year 2000 have focused on economic and financial issues (even when referring to education). The stance taken in the discussion is then more a stance usually found among managers and economists, which could be interpreted as a move towards neo-liberalism on the part of mainstream politicians.

As such interpretation of such data needs interdisciplinary knowledge and this is true since the first major work published in the field (Kucera and Francis 1967) who subjected their data to a variety of computational analyses, combining elements of linguistics, language teaching, psychology, statistics, and sociology. This

interdisciplinary analysis can be carried out with a framework which integrates findings at the micro-level of language into the macro-level of society (societal rules, social actors involved, etc.). Therefore in our analysis of the data we will consider language as a form of social practice and will be applying Critical discourse analysis (CDA)' principles (Fairclough, 1989, 22) i.e. we will consider that linguistic data are not external to society but are, in the opposite, conditioned by other (non-linguistic) parts of society.

## 2. Corpus, Methodology and Tools

### 2.1 Data and data analysis

CDA explores the often 'opaque relationships of causality and determination between discursive practices and wider social and cultural structures, relations and processes' (Fairclough 1995, 132; Widdowson 2000), such practices arising out of and are ideologically shaped by relations of power and struggles over power (Fairclough 1995, 133); the indexicality of such relationships between discourse and society would itself be a factor securing hegemony (Fairclough 1995, *ibid.*; Ochs 1991). We will then consider in our analysis the macro context where the discourse is set i.e. its social reality, after having studied the micro level, i.e. the text itself, collocations, frequencies, etc. Therefore we will interpret the data taking into account the social reality of French politics, the French political culture and journalist practices.

The purpose of the next two sections is to show present our data and how software such as SEMY (Grzesitchak Mick 2007)<sup>2</sup> or TERMOSTAT (Drouin 2003, Lesage et al 1993) can help in finding either the semantic recurrent elements or the specificities of the data and therefore pointing at the ideological positioning of discourses in journalistic corpora.

We have carried out three different studies on the journalistic discourse during the 2007 and 2012 presidential campaigns in France, campaigns where female candidates had a chance to win whether in the primaries of their party (Aubry in 2012) or in the second round (Royal in 2007).

The first study contrasts the contextual environment of the names of the three main candidates during the 2007 presidential campaign (Bayrou, Royal, Sarkozy) in order to assess their semantic profile in the journalistic discourse and to assess whether gender could have a role to play in the word choices by journalists;

The second study follows the same pattern but focuses on the textual environment of the same of the candidates of the smaller parties (i.e. Communist party, National Front, Green party, etc.), in order to check whether we find the same patterns identified in the first study;

The third study contrasts the contextual environment of the names of the two main candidates during the 2012 socialist primaries (Aubry, Hollande) in order to check whether two candidates from the same party would be presented with different discourse strategies in the same newspapers.

Our data for the Presidential campaign 2007 (partially published in Baider 2009 and Baider & Jacquy 2010) were collected over the 12 months before the French presidential elections of May 2007 for the candidates Royal, Bayrou and Sarkozy in

---

<sup>2</sup> Developed at the ATILF at <http://www.atilf.fr/>



the newspaper *Le Monde*. We collected 330 articles (roughly 100 for each candidate) focusing on one of the candidate, i.e. containing the name of the candidate in the title and several times in the text. The sum of the words was comparable for each candidate since we had roughly a total of 19.000 words for each candidate i.e. 60.000 words<sup>3</sup>. This methodology and size of data are compatible and comparable with the previous studies in language, gender and politics to which we referred in the section 1.2. Indeed these studies also used small corpora (Barnes & Larrivée worked roughly on 32 000 words; Fracchiola focused on one debate since her aim is a qualitative study of politeness strategies)<sup>4</sup>. Moreover Barnes & Larrivée used the same methodology as we did (i.e. they started out with the name of the politician in the title and then looked at the focus of the article with the repetition of the name in the body of the text), therefore our methodology and corpora can be regarded as comparable.

Our data for the Socialist primaries 2012: we gathered twenty articles from the *Factiva* data basis, focusing on each candidate for the socialist primaries from a total of six different newspapers (national dailies: *Libération*, *Le Monde*, *Le Figaro*, *La Croix*; regional dailies *Ouest-France*, *Est-Républicain*). Articles focusing on Aubry and Hollande were chosen from periods considered as the most decisive for the socialist primaries i.e. March-May 2011 is the period before the Strauss-Kahn fiasco<sup>5</sup>; June-July 2011 is the period right after the news DSK would not run; September-October 2011 is the time of the socialist primaries. Each folder for each politician exists in two versions: the first version gathers per newspaper all the articles for each candidate; the other folder is only text without details on the origin of the article. We have a total of 120 articles for Aubry and Hollande.

In previous research it was suggested that considering those discourse practices, i.e. focusing in particular on noun phrases (NPs) referring to female and male politicians, could be a way of assessing the ‘linguistic profile’ of politicians in journalistic discourse and in particular their agency (Eckert & Mcconnel-Ginet 2003). Agency is fundamental in discourse, since to be elected one could assume that to be perceived and constructed as a leader (associations with dynamism, initiative, courage, etc.) means to be constructed and perceived as an Agent, in control of the environment. This semantic value has also a pragmatic outcome: the described entity will more likely be perceived as a leader and therefore as the candidate to be elected. In order to assess the lexical environment of each NP understudy, we suggest using tools which can process a considerable amount of data, allowing quickly the identification of lexical choices and therefore help tremendously with a quantitative and a qualitative reading.

## 2.2 Methodology and Tools

Frequency and specificity are the most used criteria for corpus analysis, as previous studies on ideology and discourse have noted (Marchand & Ratinaud 2012, Michard and Ribery 2008, Valette and Rastier 2006). Especially lexical units such as

---

<sup>3</sup> At the time we did not have access to databases such as *Factiva* or *Europress*  
<http://gethelp.library.upenn.edu/guides/business/factiva.html>

See as well a more detailed data and results description in (Baider 2010).

<sup>4</sup> Cotter does not specify the size of her corpus but mentions a ‘small corpus of stories’ (2520, 2011); O’Grady does not specify either the size of the data.

<sup>5</sup> Dominique Strauss-Kahn (known also as DSK), the likely winner of the presidential elections for the socialist party, was involved in a sexual assault in May 2011 and therefore could not run in the socialist primaries.

derogative adverbs and negative adjectives were thought as the most relevant to ideological loading (Reutenauer *et al.* 2010 ; Valette 2010) and were more likely to be found when referring to “dominated entity”. As far as methodology is concerned, most recent research using corpus linguistic methodology and working on the 2012 presidential campaign such as (Marchand & Ratinaud 2012) have suggested that discursive strategies can be understood through an investigation based on semi-automatic reading.

Indeed in recent years, several software have been made available to extract lexical and syntactic patterns characteristics of the data under investigation. Gender specificities were identified applying these several software providing statistical concordance functions (ANTCONC), term extraction (TERMOSTAT) or extraction of semantic fields (SEMY). We chose among the most readily available software and free of charge<sup>6</sup>. They are also the best known by the author.

*ANTCONC* : *ANTCONC*<sup>7</sup> provides concordances, collocations, frequency lists, word lists and Key word in context (i.e. the line in which the word had been used). If concordancers are plenty, *ANTCONC* is user-friendly and quickly downloaded from the Internet onto the computer. Most importantly for researchers working on different alphabets, the software supports many different languages such as languages using Cyrillic (Greek, Russian) as well as Chinese, Japanese, Korean, etc.

*ANTCONC* enabled us to carry out a qualitative analysis since we can isolate blocks of text where the NPs are present<sup>8</sup> in order to visualize the contexts of use of any key word as in the Table 1 below (which shows what is called Key word in context or KWIC). A manual and selective reading is then possible.



Graph 1 : KWIC abstract for Aubry’s name in *le Monde*

The software also extracts and displays collocations, frequencies and word list. In particular the categories <collocates> and <cluster> allowed us to check whether some NP are more often associated than other NPs with evaluative lexical units (Michard & Ribéry and Valette’s hypothesis). Quickly we can verify a higher number of adverbs, adjectives and negative modifiers such as *un peu*, *quelques*, *moins* collocated with NP under investigation. Moreover a manual reading can point out with which *syntactic processes* these NPs are associated the most (Subject of a sentence for instance) and the agency of a unit (among other thematic roles, cf. Siewierska 1991). Therefore, with these abstracts we can observe whether NPs are constructed in discourse as a non-animated entity or at least as less in control of the

<sup>6</sup> There are many tools used in corpus linguistics. Among the free engines we can mention *Iramuteq* (University of Toulouse) <http://iramuteq.org/> for French, *Lexico 3* (University of Paris3) <http://www.tal.univ-paris3.fr/lexico/> for French and English. Among tools for copus linguistics which are not free, Sketch engine is according to the author of this paper one of the most complete and it supports also most languages: <http://www.sketchengine.co.uk/>

<sup>7</sup> <http://www.antlab.sci.waseda.ac.jp/software.html>

<sup>8</sup> <http://www.antlab.sci.waseda.ac.jp/software.html>



verb (if they are they most of the time Subjects of a verb but non – Agents for instance).

*TERMOSTAT*: TERMOSTAT (Lesage *et al.* 1993)<sup>9</sup> identifies such grammatical categories and we can quickly see which nouns, adjective or adverbs are the most used or the most specific in the data. Indeed the software presents features complementing ANTCONC's but it does not support fully languages other than French and English. The TERMOSTAT platform extracts from data made of simple and complex lexical units their frequency and specificity relatively to a reference corpus integrated into the platform 28 million of occurrences from *Le Monde* (2002) integrated to the Internet platform (Drouin 2003)<sup>10</sup> and which gathers about 8 million words. We then get a list with the most salient terms in our data, whether because they are the most frequent or whether they are the most specific to the discourse. TERMOSTAT is based on research results obtained by Lebart and Salem (1994), the criteria used to quantify the specificity of a term being the *test-value*, which is a standardized view of the frequency of the lexical units (Drouin 2003, 3). We used TERMOSTAT to examine the most specific lexical categories for each NP investigated (i.e. Aubry, Hollande), criteria which are the basis of our hypothesis. We can then decide which categories are the most relevant to a qualitative reading. The software could also be useful for teaching French as a second language since it provides with the most current vocabulary in a specific domain. In the table 2 below we can observe the most specific lexical units for the name Hollande, when compared with the reference corpus of TERMOSTAT:

Candidate	Frequency	Specificity	Spellings	Morphosyntax categ.
primaire	248	215.29	primaire__primaires	N
primaire socialiste	68	159.96	primaire socialiste	NAdj
strauss-kahniens	35	113.96	strauss-kahniens	N
candidature	153	67.28	candidature__candidatures	N
président normal	10	58.7	président normal	NAdj
favori	55	56.29	favori__favoris	N
ex-premier secrétaire	9	52.5	ex-premier secrétaire	NN
de hollande	8	51.81	de hollande	N
candidat	290	48.34	candidat__candidate	N
candidat normal	7	48	candidat normal	NAdj
projet socialiste	15	44.6	projet socialiste	NAdj
investiture socialiste	8	44.14	investiture socialiste	NAdj
présidence normale	6	43.87	présidence normale	NAdj
sympathisants	42	41.83	sympathisants	N
strauss-kahnien	6	40.6	strauss-kahnien	N
secrétaire	170	39.66	secrétaire__secrétaires	N
candidat déclaré	13	39.42	candidat déclaré	NPPAdj
présidentiable	13	39.42	présidentiable	N
humour corrézien	5	39.32	humour corrézien	NAdj
député	150	39.11	député__députés	N
sondage	97	37.69	sondage__sondages	N
candidat de substitution	4	34.19	candidat de substitution	N

**Table 1**  
Result with TERMOSTAT: NP *Hollande* in March-July 2011, in six newspapers

If linguistic expressions such as *primaire socialiste* ‘socialist primaries’ and *investiture* ‘nomination, investiture’ are expressions to be expected as specificities of a political text; however we need some knowledge about the candidate profile to

<sup>9</sup> [http://olst.ling.umontreal.ca/~drouin/termostat\\_web](http://olst.ling.umontreal.ca/~drouin/termostat_web) or [http://termostat.ling.umontreal.ca/doc\\_termostat/doc\\_termostat\\_en.html](http://termostat.ling.umontreal.ca/doc_termostat/doc_termostat_en.html).

<sup>10</sup> See Lebart & Salem (1994) for explanation on the two criteria used to quantify the specificity of a term, one of them being the “standardized view of the frequency of the lexical units”.









understand expressions such as *président normal* and *humour corrézien*. Indeed both are topoi characterizing Hollande who had the axis of ‘normality’ to organize his campaign and whose humor was well known as well as the fact that he was responsible for the Corrèze district. However, even if one does not know anything of the candidate, the incongruity of both linguistic expressions and the fact that they are specificities of the candidate profile will trigger the attention of the researcher.

As we can see with Table 2 as well other information relating to the linguistic profile of the data are provided by TERMOSTAT such as<sup>11</sup>:

- the various ‘physical appearances’ of this expression in the corpus (plural or feminine forms for nouns for instance);
- the grammatical category of the lexical unit;

These features allow a thorough quantitative reading of a corpus and enables comparison between corpora.

*SEMY* : SEMY is one of the rare software which works from a semantic based program and which is applicable to the French language. It is unfortunately only available at the ATILF. Working with semantic fields is important and SEMY aims to give a semi-automatic semantic profile of texts or linguistic entities. Indeed when fed a text or a noun, the platform extracts automatically a lexicon based on the digitalized *Trésor de la langue française* and gives the semantic field of each word in the shape of a list which Rastier would call a SEMEME.<sup>12</sup> The list provided is very similar to the one provided by TERMOSTAT, however it describes semic features and not lexical units. Indeed on the basis of the textual files provided, SEMY will give lists based on the semantic recurrence in each file i.e. using the most salient semantic feature in the data. The researcher has then the task among all the units presented to make sense of them i.e. to find some correlation in order to get a semantic profile. To do so, we group together the semic features found in the list using the CRISCO synonym dictionary<sup>13</sup>. For instance, The verb *communiquer* ‘to communicate’ was found as a recurrent semic feature in Royal’s SEMY. We look into CRISCO and find a series of verbs which are in the same semantic field as *communiquer* such as *montrer* ‘to show’ or *déclarer* ‘to declare’. In turn some of these verbs are also found Royal’s list given by SEMY, we can then organize the data accordingly, in the same semantic field. These three verbs were for instance very much present in Royal data, and much less so in Sarkozy data. This grouping, we suggest, allows us to describe the semantic profile of the NP Royal and Sarkozy in the journalistic discourse under study.

<u>montrer</u>	
<u>indiquer</u>	
<u>exposer</u>	
<u>signaler</u>	
<u>dire</u>	
<u>afficher</u>	
<u>déclarer</u>	
<u>découvrir</u>	

<sup>11</sup> Other features include as well the most frequent compound; the most frequent morpho-syntactic pattern; etc.  
[http://termostat.ling.umontreal.ca/doc\\_termostat/doc\\_termostat\\_en.html#inscription](http://termostat.ling.umontreal.ca/doc_termostat/doc_termostat_en.html#inscription)

<sup>12</sup> <http://www.atilf.fr/spip.php?article894>. SEMY was the result of a doctoral thesis and its patent is pending. For further details please contact [evelyne.jacquey@atilf.fr](mailto:evelyne.jacquey@atilf.fr)

<sup>13</sup> <http://www.crisco.unicaen.fr/des/>

annoncer	■
manifestester	■

Graph 2 Most frequent synonyms for the verb *communiquer* ‘to communicate’

In the next section we report the data found with each software and discuss our results.

### 3. Report and discussion of the results.

#### 3.1 *The semantic profiling of a noun, gender stereotyping?*

In order to work on the semantic profile of an entity in a text, we worked with SEMY and the 2007 data, since the stakes were very high, Royal being the first woman to be in the 1st presidential round and competing against Sarkozy.

During the campaign Royal complained about unfair treatment by the Media who would portray her failures or gaffes much more often than Sarkozy’s for instance. Indeed the newspaper *Le Monde* apologised for focusing on her ‘blunders’ (Baider 2009).

As mentioned in our section 2.3 under SEMY and regarding the name Royal, the most salient dimension of the candidate personality was the relational aspect: after having grouped the semic features by using the CRISCO dictionary, we obtained the Table 3 below:

CANDIDATE NAMES	SEMANTIC FIELDS	RECURRENT SEMIC FEATURES IN CORPUS
Nicolas Sarkozy	Planning	viser “to aim”, chercher “to seek”, poser “to set up”, vouloir “to want”, envisager “to envision”, compter “to reckon
	Argumentation	Entrettenir “to converse”, contredire “to contradict”, jouer “to play (on words)”
	Action	signer “to sign”, embaucher “to hire”, agir “to act”agir, embaucher “to hire”, partir “to leave”
Ségolène Royal	Political campaign	Voter “to vote”, préférer “to prefer”, élire “to elect”, désigner “to appoint”,
	Communication	Appeler “to call”, déclarer “to declare”, annoncer “to announce”, exprimer “to express”, proposer “to suggest”, téléviser “to televise”
	Defensive	Inquiéter “to worry”, renoncer “to giveup”, défendre “to defend”
François Bayrou	Debate	estimer “to evaluate”, être favorable “to be in favour”, être contre “to be against”, dénoncer “to denounce”
	Progress	progresser “to progress”, conforter “to consolidate”, poursuivre “to pursue”, proposer “to suggest” and gagner “to win”)

Table 2 Semantic fields with SEMY (Baider & Jacquey 2009)

50% of the features listed by SEMY in the context of the name Royal belonged to the field of COMMUNICATION (such as *to call, to declare, to announce, to express, to suggest*) or to the field of POLITICAL CAMPAIGN (*to vote, to prefer, to elect* and *to appoint*). In the case of her male opponents the semantic profiling was quite different. Regarding the name Sarkozy, the most salient dimension was the field of ACTION (*to sign, to hire, to act, to do, to depart*), and the field of OBJECTIVES (*to aim, to search, to covet, to want, to consider, to reckon, to wish*). The semantic profile of the name Bayrou emphasizes as well ACTION towards a goal, but in a more positive fashion than in the case of the name Sarkozy (*to progress, to consolidate, to continue, to want, to offer, to make*); it emphasized also REBELLION (*to be unfavorable, to be against, to denounce, to disapprove, etc.*). If we compare the three semantic profiles, we identify

the gender stereotype defining the female candidate and the male candidates to be defined in an opposition ACTION vs. COMMUNICATION. This opposition reminds us of the cliché of women being good in relating to people and men being better in taking steps and making decisions.

These results are explained either by the discourse of the candidates themselves or their behavior (Royal and Sarkozy both played on sexual stereotypes) or by the perception the journalists have of the politicians. We cannot determine the source of the difference but the difference is there.

From a semantic point of view, SEMY has then revealed what we call “the deep structure” of the meaning of the names Bayrou, Royal, Sarkozy as constructed in discourse. However the software TERMOSTAT (Drouin 2003, Lesage *et al.* 1993) reveals another facet of the meaning with revealing the *specificities* of words found in the context of each name and confirm a difference for each name. This is the topic of the next section.

### 3.2 *Antconc, KWIC abstracts and agency: power in politics?*

With the concordance ANTCONC we worked on KWIC abstracts such as the ones listed in the section 1.3. To evaluate agency, we focused on the syntactic functions of the names. Eckert and McConnell-Ginet pointed out that any NP in the position of subject in a sentence will be associated with power and control (2003: 52). We also evaluated the semantic functions, in order to examine the actual *causality* of each NP, and restricted our study to Agent and non-Agent categories (Dowty 1991; Siewierska 1991). This evaluation has to be done manually (Baider 2009.) We found that the syntactic choices made in journalistic discourse when referring to the presidential candidates supported a partial dichotomy with the nouns.

	Sarkozy		Royal	
<b>Subject</b>	158 (21)	63.00 % (13.30%)	135 (48)	46.55% (35.60%)
<b>(Verb copula) (Passive voice) (Unaccusative Verb)</b>				
<b>Object (direct)</b>	15	5.60%	14	4.83%
<b>SN (<i>de</i> + Sarkozy or Royal)</b>	37	14.75%	68	23.45%
<b>SP</b>	41	16.33%	73	25.17%

Table 3. Syntactic roles for NPs Royal and Sarkozy drawn from KWIC abstracts (Baider & Jacquey 2010)

Sarkozy functions more often as a Subject as the noun Royal. This contrastive study of syntactic choices used to describe a female and a male candidate confirms a perception of the journalists conforming to stereotypes: the noun of the male candidate is, syntactically, in control of the verb, while the noun of the candidate is its subordinate. A difference was even greater as far as the semantic functions were concerned.

	Sarkozy	Royal
	251	290

Non-agent = Theme, experiencer, location, cause, beneficiary	114	= 45.41%	212	= 73,10 %
Agent	133	= <b>52,98%</b>	74	= <b>25,51%</b>
?	4	= 1,50%	4	= 1,40%
		(99.93%)		(99.98%)

Table 4. Thematic role of the noun *Sarkozy* (251 verbs) and the noun *Royal* (290 verbs) (Baider 2009)

The male candidate is constructed first as an ‘Agent’ which confirms the findings with SEMY. On the contrary, the name of the female candidate played first the role of an aim (*for* Royal), a place (*chez* Royal) or a relation (*de* Royal) which we gathered in the category ‘Cause’. Moreover, the noun *Royal* was described twice as much as *Sarkozy* as an ‘Experiencer’, i.e. as a person who feels or a person who does not have the control on the action described in the sentence (Baider 2009). As a matter of fact, if, in the same sentence both names of the candidate are found, the noun *Royal* will be found in 75% of the cases in a clause embedded to the clause in which the noun *Sarkozy* is the subject, as the occurrences found in corpus show below (Baider 2009):

(1) Prenant de vitesse *Ségolène Royal*, qui doit débiter, lundi 25 septembre, un voyage au Sénégal, *Nicolas Sarkozy* a effectué, l'avant-veille, un déplacement de quelques heures à Dakar ;  
‘Taking by surprise *Ségolène Royal* who will start on Monday the 25th of September a journey to Senegal, *Nicolas Sarkozy* did, the day before, a trip of a few hours to Dakar’.

(2) *Nicolas Sarkozy*, qui battra *Mme Royal* 36 % contre 34 %, (...)  
‘*Nicolas Sarkozy*, who would beat Ms *Royal* 36 % against 34 %, (...)’

(3) *M. Sarkozy* croit que *Mme Royal* « le protège beaucoup ».  
‘*M. Sarkozy* believes that Ms *Royal* “protects him a lot”’.

The opposite is less likely to be found:

(4) Au moment où *Nicolas Sarkozy* s'apprête à rendre publique sa candidature (...) *Ségolène Royal* prend le large.  
‘When *Nicolas Sarkozy* gets ready to announce publicly his candidacy (...), *Ségolène Royal* takes a trip’

We can also observe the semantic prosody (Bednarek 2008) in all the occurrences above as being consistently negative for *Royal*: *Royal* is surprised by *Sarkozy* quick move; whereas he announces his candidacy (an important move), she goes on a trip (a trivial endeavor), the female candidate being as well the target of ironic observations (3) and (4).

The French press reports on *Royal* during the 2007 presidential campaign may have been driven by and perpetuated a normative view of the sexes. Indeed explanation for these results could be as much the responsibility of the politician herself and her/ his choices of words and personae. Therefore their linguistic profile could also be considered as an exception but not the rule given our other results presented in the next section.

#### 4. Smaller candidates: specificity, power and gender in politics

In order to test the validity of our previous hypothesis whether discourse in newspapers testify to gender or power relationships in politics, we analyzed smaller



candidates with powerful female candidacy and their linguistic profile in order to test whether we would find similar results i.e. gender relevance would appear when power is at stake:

- In the 2007 presidential elections, we assessed the linguistic profiles of smaller candidates such as Voynet, Le Pen,
- In the 2011 socialist primaries we analysed the specificities of the data focusing on Aubry and Hollande who belong to the same party.

#### 4.1 *Qualitative reading, Semy and Antconc, the agency criteria*

In a previous study (Baider & Jacquey 2010) we could not find any of the asymmetry observed for Royal, and Sarkozy with another corpus gathering on the smaller candidates during the 2007 campaign. Indeed using the same methodology we identified semantic profiles and agency for candidates of smaller parties such as Voynet for the Greens, Buffet for the communist party and Besancenot for the radical left, all candidates who had no chance to win.

The lexical asymmetry found in relation to Royal did not hold true either for the other female contenders who are part of our study. For example, for the male and female candidates in general, *negativity* was more a masculine trait than a feminine one, whereas in the case of Royal *negativity* was typical of the female candidate. Indeed, the noun *Bové* was defined by the field of **Unlawfulness**; *Besancenot* by the field of **Discord** and *Le Pen* by the field of **Concealment**. The three female politicians were characterised, as was Royal, by their relational personalities with the field of **Advice** for *Buffet*, of **Moderation** for *Voynet* and of **Persuasion** for *Laguiller*.

As far as grammatical choices are concerned the female Communist Party leader Buffet and her newly appointed male opponent, Besancenot, the Revolutionary Communist party representative, served as a case in point. Both names were considered together, since the NP *Buffet* and *Besancenot* are found in the same syntagm 89% of the time when they are coordinated with another NP (such as in *Buffet et Besancenot*). At almost all levels of discourse our results contrast with the figures obtained with the NPs Bayrou, Sarkozy and Royal. Indeed:

- The name *Besancenot* is used with another NP and therefore 'not autonomous' in 54% of the occurrences, whereas the name *Buffet* stands alone 35% of the times.
- The female candidate, Buffet, is twice as often an agent as her young male counterpart.
- On the lexical level, Besancenot, a Trotskyite, is also described as Alain Krivine's "young protégé" (*jeune poulain*), Krivine being a well-known and longtime figure of the extreme left. The term 'protégé' reinforces his (only) burgeoning career. By contrast, the name Buffet co-occurs with the adjective *fort* (strong) and the noun *chef* (leader) signaling the long political career of the communist leader.

Even when both names occur in the same sentence, there is no striking pattern of hierarchy to be noticed: both names are used in the same clause, both function as a subject or as part of a prepositional phrase.

The criteria above are the only ones making a difference between both discourses and they all play in favour of the females candidates. We may witness a case here that when the stakes are not as high as a presidential election, reproducing gender stereotypes in the political sphere did not seem to be on the stake holders' agenda, i.e. Buffet (the then communist party leader) was granted more agency than her young opponent Besancenot (the then extreme left leader). Parity can be afforded when there

is not much to lose or to gain<sup>14</sup> as Moores had already observed in 1997 in her work on women in French politics. However as mentioned before (Baider 2009, Baider & Jacquy 2010) the Sarkozy and Royal candidacies were well-known to be a somewhat exceptional case, since Royal and Sarkozy had played for the first time in the French political history on sexual stereotypes (Gingras 1997; Achin & Dorlin 2008; Coulomb-Gully 2012).

#### 4.2 Quantitative reading, Termostat, the specificity criteria

We described the data previously but we would like to point out again that we are working on the socialist primaries in which one female and one male candidate were the final contenders. However they were not the favorites from the beginning since Dominique Strauss Khan (DSK) was clearly the favorite of the party. However being caught in a scandal, he had to withdraw from the race by the end of May.

##### 4.2.1 Before the DSK affair

If we look at the Table 5 below, the most frequent terms provided by TERMOSTAT will be the terms found in articles focused on political campaign (*candidate, first round, vote, campaign, party nomination, meeting, etc.*). They are therefore similar for all candidates.<sup>15</sup>

Hollande		Aubry	
Items	Frequency	Items	Frequency
candidat	290	candidat	300
primaire	248	secrétaire	234
président	171	projet	212
secrétaire	170	primaire	204
candidature	153	parti	161
député	150	candidature	157
campagne	121	socialiste	109
parti	108	sondage	94
sondage	97	mai	90
soutien	83	juin	85
projet	82	maire	83
maire	72	député	82

Table 5: The most frequent items around the NPs *Hollande* and *Aubry* in March-June 2011 in the six newspapers (Baider & Jacquy 2014)

However when looking at the specificity criteria and before the Strauss Khan fiasco, we observed the name Aubry to be described extremely positively. Actually after a close reading of newspapers discourse with TERMOSTAT, *no* blatant gender stereotyping could be witnessed in the six newspapers we took for building our corpus. In the opposite, we observed reverse stereotypes. The discourse described the name Aubry as a dynamic and organised manager, and the name Hollande as an embattled politician busy dealing with conflict. In the table 6 below listing the specificities: three odd expressions typify the candidates (i.e. when to take away all the expressions common to both candidates).

<sup>14</sup> A tactic which is also followed by some parties when drawing their list of candidates for elections: women are present in region where stakes are not high or where they are most likely to lose (Sineau 2010).

<sup>15</sup> TERMOSTAT would be useful for teaching French as a foreign language in order to spot the most current lexical fields when describing a certain event.

In order to categorize the lexical units, first we used the interface of the morpho-syntactic and semantic criteria (denoting positive or negative values) i.e.:

- verbs (denoting then action) are underlined,
- adjectives (describing positive or negative assessment) in italics,
- nouns (with a positive or negative connotation) in bold

We have categorized the lexical units categorized by their morphosyntactic categories.

Verbs are only found with the NP Aubry and not with Hollande which can denote a more active and dynamic personality (or at least of a perception of a more dynamic person):

Aubry: MA says (*dit Martine Aubry*), MA does (*fait Martine Aubry*), MA replies (*réplique Martine Aubry*).

Nouns referring to the candidates are more positive for Aubry than for Hollande who is referred to as a challenger, as a substitute and the expression except Hollande is found in *tout sauf Hollande* (all except Hollande):

Hollande: depute (député), president of the council , challenger, substitute, (candidature de substitution), Neerland (Hollandais), except Hollande

Aubry: manager (*patronne*), rallying figure (*rassembleuse*)

Adjectives are found describing Hollande but not Aubry; they are not favorable (*anti-hollande*) or they describe the theme of his campaign that he will be a ‘normal candidate’; *normal* as adjective can be interpreted as a choice to describe Sarkozy, his political opponent, as not being a ‘normal’ president (because of his flashy life-style, as well as being far from ordinary citizens) or to describe Aubry as not being a usual candidate (because of her gender).

Hollande: normal (*président, présidence, candidature*), corrézien, anti-hollande,

Hollande		Aubry	
N-grams	Spec	N-grams	Spec
<i>primaire socialiste</i>	155.97	<i>primaire socialiste</i>	122.00
<i>strauss-khanien(s)</i>	113.69	<b>de Martine Aubry</b>	69.52
<b>Candidature</b>	67.11	<i>projet socialiste</i>	64.95
<i>président normal</i>	47.89	<i>strauss-khanien(s)</i>	60.24
<i>projet socialiste</i>	44.49	<b>Patronne</b>	49.52
<i>ex-premier secrétaire</i>	52.38	<b>emplois d’avenir</b>	45.02
<i>candidat normal</i>	47.89	<i>conseil politique</i>	35.93
<i>projet socialiste</i>	44.49	<i>après DSK</i>	35.08
<i>investiture socialiste</i>	44.03	<u>que Martine Aubry</u>	35.08
<i>présidence normale</i>	43.77	<u>dit Martine Aubry</u>	35.08
<b>Sympathisants</b>	41.72	<b>Rassembleuse</b>	35.08
<i>candidat déclaré</i>	39.32	<i>Présidentiable</i>	34.89
<i>humour corrézien</i>	38.73	<i>égalité réelle</i>	34.06
<b>Député</b>	38.73	<i>parti socialiste</i>	33.92
<b>candidat de substitution</b>	34.11	<i>bureau national</i>	32.79
<b>Hollandais</b>	32.54	<i>candidat déclaré</i>	31.99
<i>position de favori</i>	29.17	<b>primaire de désignation</b>	28.93
<b>Biceps</b>	28.54	<i>soubassement intellectuel</i>	28.93
<i>pro-dsk</i>	28.13	<i>convention nationale</i>	26.64
<b>sauf Hollande</b>	28.13	<b>emplois jeunes</b>	25.15
<b>Présidentialité</b>	28.13	<b>délai de décence</b>	25.03
<i>front anti-hollande</i>	28.13	<i>projet adopté</i>	24.74
<i>expérience ministérielle</i>	27.81	<b>texte de soutien</b>	23.31
<b>Ralliement</b>	25.82	<b>coup de tonnerre</b>	22.75
<i>contrat de génération</i>	24.33	<i>jeunes socialistes</i>	22.40
<i>Rocardiens</i>	24.33	<b>Partisans</b>	21.68
<i>stature présidentielle</i>	24.33	<b>primaire d’octobre</b>	21.25
<i>dépôt officiel</i>	24.33	<u>Martine Aubry fait</u>	21.25

<i>opposant farouche</i>	24.33	<u>réplique Aubry</u>	21.25
<i>justice fiscal</i>	24.04	<b>face à face</b>	21.25
<i>referendum européen</i>	24.04	<u>Martine Aubry a</u>	21.25
<b>Soutien</b>	23.29	<i>thèmes économiques</i>	21.25
<b>président de conseil général</b>	22.68	<b>coude à coude</b>	21.25
<b>Challenger</b>	22.65	<i>image de rassembleuse</i>	21.25
<i>adversaire socialiste</i>	21.79	<i>candidature potentielle</i>	21.25

Table 6. The most specific n-grams associated with the NPs *Hollande* and *Aubry* in March-July 2011 in the six newspapers (Baider & Jacquy 2014)

We can also observe that the general tone of the specificities for Hollande are more negative than for Aubry (*adversaire*, *opposant farouche*, *sauf Hollande*, *front anti-Hollande*), an hostility which we have not found in Aubry specificities.

Our findings are corroborated by other studies. Indeed Marchand & Ratinaud (2012, 692) list the most specific words as found in the socialist candidates *speeches*. For Aubry out of 35 specific words, almost 40% are verbs, half of them denoting dynamism and actions such as *battre* ‘to beat’, *augmenter* ‘to rise’, *arrêter* ‘to stop’, *défendre* ‘to defend’, *fermer* ‘shut down’; in contrast only 12 % of the most specific words in Hollande speeches are verbs and they do not denote actions (*falloir* ‘must’, *avoir* ‘to have’, *évoquer* ‘to refer’). These topoi described in 2007 Sarkozy and Royal respectively.

#### 4.2.2 After the DSK affair

However we noted the shift in the linguistic portraits during the summer 2011, once it was clear that both Aubry and Hollande were favourites for representing the party against the incumbent president. Despite his lack of personality and the hostility he generated as described in the section above, Hollande seemed to have become the consensus of the left. Indeed, when three months before Aubry was the ‘*rassembleuse*’ (see Table 6 above), we found in the August-September data that *vote Hollande c’est pour éviter la fragmentation* ‘voting for Hollande will avoid fragmenting’, *il est mieux à même de rassembler la gauche* ‘he is better suited (than Aubry) to unite the left’. We can then guess a reversal in the party discourse reflected in the journalist word choices: Hollande was thought more ‘apt’ to beat Sarkozy.

Why would Hollande be better suited for the job than Aubry? Maybe this choice was partly explained by the lead Hollande had in the opinion polls, a lead which is in itself hard to understand since many abstracts show all along our 2012 study (Baider & Jacquy 2014). We could only conclude that, despite a positive ethos, Aubry never was the favourite either in the polls, either in the party. By contrast, despite a catastrophic portrait, Hollande lead in the polls, after Strauss Khan withdrawing from the race and his party supported him.

We could then conclude from the data given in this paper for the 2007 and 2012 campaign that female and male politicians could be treated on the same footing by their party, and / or in the French press as long as the stakes are not high; it seems that once they get higher such as being potentially elected president, female politicians face negative discrimination. One could say of course that journalist being close to political parties these findings could be just a way of manipulating people, so that the preferred candidate by the party is chosen by the electorate. However as most studies in French political life have shown (Sineau *et passim*; Coulomb-Gully 2012; Bertini

2007, Gerber 2009, Adler 1993) the ‘chosen candidate’ in French politics is a gender issue, whether in the party and whether by the public.

### **Concluding remarks**

In concluding this study, extraction of discourses markers (mainly lexical and syntactic) using NLP technologies can address the issue of gender or political favoritism in journalistic discourse. We showed that conclusions drawn from qualitative studies may be supplemented, refined and nuanced by working the same data using specific computer software. We have seen as well that SEMY based on definitions of the digital version of *Le Trésor de la langue française* allowed a reading taking into account the most salient semantic features found in the contextual environment of the names Royal, Bayrou and Sarkozy. Then the linguistic profile for each candidate found with SEMY has been clarified and supplemented by the use of TERMOSTAT which works with specificities.

Even though these observational tools must always be improved, since the lists they provide require manual work, we can only recommend adding these tools when working on textual meaning, provided that the researcher has a good knowledge of the context in which the data have been produced. We would then conclude with the three following recommendations:

- Manual studies are essential for syntactic-semantic research since for the moment they are impossible automatically ; such qualitative reading also allow the researchers to know very well their data, a prerequisite for deciding on which interpretation would be the best suited. ANTCOnc provides the KWIC abstracts which facilitate greatly this evaluation;
- SEMY reveals recurring semantic features which cannot be detected with a superficial reading: it brings to light isotopies (Rastier 1987) which are the semantic characteristics recurrent in a text;
- TERMOSTAT reveals how these isotopies are realized and how different they are for each candidate.

The three methodologies therefore appear complementary to the analysis of textual meaning. However to draw broader conclusions from these data and as far as gender is concerned, we have to examine the relationships between the other dimensions of the communicative event which is journalistic text, as Schiffrin (2006) and Cotter (2011) underscore. They are ethnographically situated pragmatic arrangements and must obey constraints, sociocultural values, and community-of-practice conventions (Cotter, 2011: 2532). In our case discourse practices (i.e. here the processes of text production in the French press) and socio-cultural practices (i.e. here the French society) should be taken into account in linguistic investigations of social power, since we focus on the “connection between words and the world” (Schiffrin, 2006:333). The combined discussion of quantitative and qualitative analyses and the contexts of use (particularly social attitudes towards gender) put into light the relations between “practitioner, text, and the shared world of all participants in the discourse; and what that can say about the role of women at any point in time in a particular sociocultural context” (Cotter, 2011:2532). We then recommend the use of corpus linguistic tools as a valuable method in political and social science and they could be a prerequisite for qualitative research in digital humanities.

### **References**



Achin, Catherine & E. Dorlin (2008). Nicolas Sarkozy ou la masculinité mascarade du président, *Raisons politiques* (31) 19-46.

Adler, Laure (1993). *Les femmes politiques*. Paris: Seuil.

Alim Samy & Angela Reyes (2011) Introduction: Complicating race: Articulating race across multiple social dimensions, *Discourse Society* 22: 379-384.

Baider Fabienne & Evelyne Jacquey (2010). Substantive Dis-Embodiement, Syntactic Embedment. *Feminism, Femininity and Gendered Discourse*, (Eds. J. Holmes & M. Marra). Cambridge Scholar Press. 145-167.

Baider Fabienne. (2009). Lexical change, discourse practices and the French press: Plus ça change, plus c'est la même chose? *Language Variation – European Perspectives II*, (Eds. S. Tsiplakou, M. Karyolemu et P. Pavlou), 27–46. John Benjamins: Amsterdam.

Baider Fabienne & Evelyne Jacquey (2014). Data mining for Tracking Gender Bias: a Curse or a Blessing? The 2012 presidential duos. *Ecological and Data-Driven Perspectives in French Language Studies* (eds H. Tyne et al.). Newcastle upon Tyne: Cambridge Scholar Press. 222-245.

Baider Fabienne & Constantinou Maria (2014). How to make people feel good when wishing hell: Golden Dawn and National Front Discourse Discourse, Emotions and Argumentation *2014 Yearbook of Corpus Linguistics and Pragmatics*, Springer (Ed. J. Tromero). Springer.

Bertini, Marie.-Josèphe. (2002) *Femmes, le pouvoir impossible*, Paris, Pauvert/Fayard.

Brasart, Patrick (1994) Petites phrases et grands discours. *Mots. Les langages du politique* 40 : 106-112.

Caillat Dominique (in print). Le recours au discours rapporté dans le débat présidentiel de l'entre-deux tours de 2007 : des usages sexués? *Les discours rapportés dans les discours médiatiques, une affaire de genre ?* Eds F. Sullet-Nylander, M. Roitman, L. Rosier, S. Marnette et J.-M. Lopez. Stockholm: Stockholm University Press.

Clark Kate (2005 [1998]). The linguistics of blame. *The Feminist critique of Language* (ed. D. Cameron). Routledge.

Cotter C. (2011) Women's place at the Fourth Estate: Constraints on voice, text, and topic. *Journal of Pragmatics* 43: 2519–2533

Coulomb-Gully, Marlène (2009). Présidentielles 2007 : Média genre et politique. *Mots* 90 (1) 5-11.

Coulomb-Gully, Marlène (2012). *Présidente : le grand défi !* Éditions du Seuil, Paris.

Devriendt Emilie (in print). Du discours rapporté dans deux séquences médiatico-discursives de la pré-campagne présidentielle française 2012: candidats vs. candidates, une loi du genre? *Les discours rapportés dans les discours médiatiques, une affaire de genre ?* Eds F. Sullet-Nylander, M. Roitman, L. Rosier, S. Marnette et J.-M. Lopez. Stockholm: Stockholm University Press.

- Drouin, Patrick (2003). Term extraction using non-technical corpora as a point of leverage. *Terminology* 9 (1) 99-117.
- Dupuy P.-O., Marchand P. (2013). Évolutions lexicales des débats de l'entre-deux-tours de l'élection présidentielle française. *Bulletin Vals-Asla (Bulletin suisse de linguistique appliquée)* 98 :105-128.
- Eckert, Penelope & Sally McConnell-Ginet (2003). *Language and gender*. Cambridge: Cambridge University Press.
- Fairclough Norman (1989). *Language and Power*. London: Longman
- Fairclough Norman (1995). *Media Discourse*. London: Edward Arnold.
- Gerber, Gwendolyn L. 2009. Status and the gender stereotyped personality traits: Toward an integration. *Sex Roles* 61: 297–316
- Gidengil Elizabeth & Everitt Joanna (2003). Talking Tough: Gender and Reported Speech in Campaign News Coverage. *Political Communication* 20, 209-32.
- Grzesitchak Mick, Evelyne Jacquy & Mathieu Valette (2007). *Systèmes complexes et analyse textuelle : Traits sémantiques et recherche d'isotopies*, ARCo'07 – Cognition, Complexité, 227-235.  
[http://mathieu.valette.free.fr/Articles/Grzesitchak\\_Jacquy\\_Valette\\_2007.pdf](http://mathieu.valette.free.fr/Articles/Grzesitchak_Jacquy_Valette_2007.pdf)
- Goodwin M. H. (2011). Engendering Children's Play: Person Reference in Children's Conflictual Interaction." In *Conversation and Gender*. S. A. Speer and E. Stokoe, eds. 250-271. Cambridge: Cambridge University Press.
- Kahn, Kim Fridkin (1994). The distorted mirror: Press coverage of women candidates for statewide office. *Journal of Politics* 56 (1) 154-173.
- . 1996. *The political consequences of being a woman: How stereotypes influence the conduct and consequences of political campaigns*. Columbia University Press: New edition.
- Kanter, Rosabeth M. (1977). *Men and women of the corporation*. New York: Basic Books.
- Kitzinger C. (2007). Is 'woman' always relevantly gendered? *Gender and Language* 1(1): 39-49.
- Kohl, Diane. (2011). The Presentation of "Self" and "Other" in Nazi Propaganda. *Psychology & Society* 7 – 26.
- Krieg Planke A. (2011). Les « petites phrases » : un objet pour l'analyse des discours politiques et médiatiques *Communication & Langages*, 168 : 23 - 41
- Kucera & Francis (1967) *Computational Analysis of Present-Day American English*.
- Lazar, Michelle M. (ed.) (2005). *Feminist critical discourse analysis: Studies in gender, power and ideology*. Palgrave Macmillan.
- Le Bart, Christian (2010). Parler en politique. *Mots. Les langages du politique* 94 : 77-84.

Lesage, René; Price, Wilson; Bissonnette, Carole & Patrick Drouin (1993). Enquête sur l'état d'utilisation des outils automatisés d'aide à la rédaction dans les organisations. *META* 38 (2) 367-389.

Macaulay, Monica & Coleen Brice (1997). Don't touch my projectile: Gender bias and stereotyping in syntactic examples. *Language* 73 (4) 798-825.

Marchand P., Ratinaud P. (2012). L'analyse de similitude appliquée aux corpus textuels : les primaires socialistes pour l'élection présidentielle française (septembre-octobre 2011). Actes des 11eme Journées internationales d'Analyse statistique des Données Textuelles (JADT), 13-15 juin 2012, p. 687-699.

Marchand P. (2009). Genre, style et attitude à l'égard du langage : tentative de diagnostic automatique sur un corpus politique. Journées d'études « Le style et ses modélisations » Université de Tours, 10-11 décembre 2009. [http://pascal-marchand.fr/IMG/pdf/PM\\_Tours.pdf](http://pascal-marchand.fr/IMG/pdf/PM_Tours.pdf)

Michard, Claire (1988). Some socio-enunciative characteristics of scientific texts concerning the sexes, the nature of the right. *A feminist analysis of order patterns*, (ed. G. Seidel), 27-61. Amsterdam: John Benjamins.

Michard, Claire & Claude Ribéry (2008 [1987]). *Sexisme et sciences humaines Pratique linguistique du rapport de sexage*. Lille: Presses Universitaires du Septentrion.

Moore, Pamela M., (1997). Women in the presidential race: Arlette Laguiller and Dominique Voynet, 1995. *Women in French Studies* 5, 281-292.

Mussolff, Andreas. (2008). What can critical metaphor analysis add to the understanding of racist ideology? Recent studies of Hitler's anti-Semitic metaphors. *Critical Approaches to Discourse Analysis across Disciplines* 2 (2):1-10.

Ochs, Elinor. (1992). Indexing Gender. *Rethinking Context : Language as an Interactive Phenomenon*. A. Duranti and C. Goodwin (eds). 335-358. Cambridge, Cambridge University Press.

Peck, Jennifer (2006). Women and promotion: The influence of communication style. In: Barrett, Mary and Marilyn Davidson (eds.) *Gender and Communication at Work*. Hampshire: Ashgate, 50-66.

Penelope, Julia (1990). *Speaking Freely: Unlearning the Lies of the Fathers' Tongues*. New-York: Pergamon Press.

Rabatel, Alain (in print), 'La parole des politiques soumise à contre-enquête' représente-t-elle de la même façon la parole des hommes et des femmes politiques ?'. *Les discours rapportés dans les discours médiatiques, une affaire de genre ?* Eds F. Sullet-Nylander, M. Roitman, L. Rosier, S. Marnette et J.-M. Lopez. Stockholm: Stockholm University Press.

Rastier, François(1987). *Sémantique interprétative*. Paris: PUF.

Reutenauer Coralie, Mathieu Valette & Evelyne Jacquey (2010). De l'annotation sémique globale à l'interprétation locale : environnement et image sémiques d'« économie réelle » dans un corpus sur la crise financière, *Cognitica – Actes du colloque de l'Association pour la Recherche Cognitive Arco '09 : Interprétation et problématiques du sens* (9-11 November 2009, Rouen), 29-39.

- Reutenauer, Coralie (2009). Analyse et modélisation sémantique à partir de ressources lexico-sémantiques : résumé, *Texto !* (XIV),1, <http://www.revue-texto.net/index.php?id=2093>.
- Roncarolo, Franca (2000). A woman at the Quirinal: Thanks, but no thanks. The social construction of women's political agenda in the 1999 Italian Presidential Election. *European Journal of Women's Studies* 7 (1) 103-126.
- Saïtta, Eugénie (2008) Les journalistes politiques et leurs sources. D'une rhétorique de l'expertise critique à une rhétorique du cynisme », *Mots. Les langages du politique* 87 : 113-128.
- Schiffrin, Deborah, 2006. *In Other Words: Variation in Reference and Narrative*. Cambridge University Press, Cambridge.
- Sinclair, John M. (1991). Words and phrases. *Corpus, Concordance, Collocation*. Oxford: OUP. 70–75.
- Sineau, Mariette (2010). *La Force du nombre : femmes et démocratie présidentielle*. Ed. de l'Aube : Paris.
- Shaw Sylvia (2013) An Ethnographic Investigation into Gender and Language in the Northern Ireland Assembly, I. Poggi et al. (eds.): *Political Speech 2010*, LNAI 7688, pp. 39–53
- Sreberny-Mohammadi, Annabelle & K. Ross (1996). Women MP's and the media representing the body politic. *Women in Politics* 49 (1) 103-117.
- Sung Matthew Chit Cheung (2013). Media Representations of Gender and Leadership: From a Discourse Perspective. *Brno Studies in English* 39 (1) DOI: 10.5817/BSE2013-1-5.
- Thimm, Caja, Sabine Koch and Sabine Schey (2003). Communicating gendered professional identity. In: Holmes, Janet and Miriam Meyerhoff (eds.) *The Handbook of Language and Gender*. Oxford: Blackwell, 528–549
- Valette, Mathieu (2010). Propositions pour une lexicologie textuelle, *Zeitschrift für Französische Sprache und Literatur* 37, 171-188.
- Valette, Mathieu & François Rastier (2006). Prévenir le racisme et la xénophobie – propositions de linguistes. *Les langues modernes* 2, 68-77.
- van Dijk, Teung. (2006). Discourse and manipulation. *Discourse & Society* 17(2): 359-383.
- Widdowson, Henry. G. (2000). On the limitations of linguistics applied. *Applied Linguistics* 21 (1): 3-5.