



**HAL**  
open science

# History and Information Technology : The French are way behind

Damon Mayaffre

► **To cite this version:**

Damon Mayaffre. History and Information Technology : The French are way behind. *Lexicometrica*, 2001, 3. hal-01364901

**HAL Id: hal-01364901**

**<https://hal.science/hal-01364901>**

Submitted on 31 Jan 2020

**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.

## Damon Mayaffre

CNRS (UMR 6039 Bases, corpus et langage)-Université de Nice

Adresse : 37 rue des tilleuls, 04120 Peyroules. France

Tel/fax : 04-93-60-31-17

E.-Mail : mayaffre@unice.fr

## History and Information Technology : The French are way behind

The trends in historical studies undergo the same see-saw movements as can be observed in the evolution of public opinion. In France, the methodical, pluri-disciplinary and very often quantitative historiography has given way to a militant revival of the mono-disciplinary, intuitive approach. In view of a number of excesses or the sterility of a few undertakings which claimed to be scientific, a new form of journalistic historiography is currently emerging.

Now is a time of “coffee-table” biographies : a case in point being the latest opus on Leon Blum, in which his various marriages “with Lise”, “with Thérèse”, “with Janot”, play a central role<sup>1</sup>.

It is also a time for what can be called “diplomats”, that naive and pseudo scientific belief that the whole of the history of man is documented in secret and hitherto untapped archives, or in the personal diaries of great men or in Cabinet papers, all of which historians are beginning to hunt down furiously, like journalists looking for a scoop.

It is a time for shying away from conceptualisation and returning to a narrow vision of the field, a far cry from the attempts of the sixties and seventies.

Finally, it is a time of immodest triumph for subjectivity or historical impressionism, just when a number of tools, such as I.T., allow for more rigorous and farther-reaching descriptions of the objects under study.

## Current political history , discourse analysis and the computer

Let us here illustrate these points in concrete terms in the specific field of the analysis of political discourse in contemporary history.

Faced with corpora that he might wish to dissect as interesting historical objects, it seems natural for the historian to start by raising the issue of textual analysis. To mention it is already too bold for contemporary historiography. Today, the only approach tolerated by the institution is the “intuitive” one, based on the premise that there is such a thing as “a natural understanding” of the language being used, which leads to the kind of literary commentary which is expected from sixth formers. This “*good old (reading) method which is anything but a method*”<sup>2</sup>, far from questioning itself tries to exert a monopoly. Any other approach, insisting on method but not excluding interpretation, even if it does not aim at replacing the traditional approach but merely at complementing it, will be suspect, open to criticism and, in the end, rejected.

---

<sup>1</sup> I. Greilsammer, *Blum*, Paris, Flammarion, 1996, 611p. The quotations are the actual headings for parts 2, 3 and 4 of the work.

<sup>2</sup> A. Prost, “Les mots”, in R. Rémond (ed.), *Pour une histoire politique*, Paris, Seuil, 1988, p. 282.

Young PhD students, researchers in history, thus find themselves in a surprising situation. Should they come before the jury armed solely with their common sense and their personal convictions to analyse the text, then they will avoid criticism and be congratulated for the subtlety of their analysis. If they are careful to define their method, to explain their procedure and to submit themselves to definite criteria, then they arouse suspicion and lay themselves open to Byzantine criticism on the part of jury members who, all of a sudden, become sensitive about the epistemology of their subject.

The method of textual analysis that we used in our doctorate work has not eluded this criticism. And if our doctoral thesis<sup>3</sup> was unanimously acclaimed by the jury it was not thanks to its rigorous method but *in spite* of it. Our merits as a historian were warmly acknowledged *notwithstanding* the methodical rigour of our work, indeed almost *despite* its scientific ambition.

And yet the method used –lexicometry– is simple. It consists in measuring the changes in the vocabulary used in political discourse, by plotting the frequency distribution of a given word in the speech of a locutor; evaluating how a term is over- or under-used as compared to the average use in right- or left-wing discourse; recording the chronological distribution of these words in a diachronic corpus to evaluate the changes in political stances. If it is methodical work it is because the historian-lexicographer defines his principle of analysis and rule of conduct; and because of this, invites criticism. He claims not to assert anything unless it is demonstrated or substantiated by some form of quantitative measurement. As for his ambitions, they are simple and limited : to delay for as long as possible the moment of interpretation, to push back his own subjectivity. Instead of dealing with a raw text, to be investigated and interpreted straight away without any mediation, the historian is now presented with a text which has already been processed, systematically sorted and indexed. Lexicometry does not preclude historical interpretations, but the descriptive data supporting those interpretations is rigorously impartial. Indeed, the interpretations might be all the more daring since they stem from such an objective description.

Thus, by this method and thanks to a synchronic approach of the speeches of Thorez, Blum, Flandin and Tardieu in the thirties we have been able to bring up to date the lexical and ideological identity of the communists, socialists, Orléanists and Bonapartists, and more generally the mentalities of the Right and the Left. Above all, thanks to a diachronic study covering the 10 crucial years from 1928 to 1939, we have been able to illustrate the verbal civil war waged during the last decade of the Third Republic. This battle of words, very often fought from changing positions (cf. the constant reshuffles between Left and Right on the question of patriotism and defence of the Republic) led to the implosion of the republican consensus on the eve of the war and allows for a better understanding of “la drôle de guerre”, the defeat and Vichy. More precisely, it has been possible to study the tremendous ground covered by the French communist ideology between 1930 and 1939, from scientific to humanist socialism, and to establish a new sequencing of the period. Similarly we established the drift of the moderate Right towards fascism or legitimism. In particular it seems obvious that the discourse of Vichy was foreshadowed, as early as the years 1935-1939, by the speeches of the Democratic Alliance – the first centre-right party – where the terms “family”, “labour”, “fatherland” and the whole of the ruralist vocabulary and ideology blossom.

---

<sup>3</sup> On what follows cf. D. Mayaffre, *Le poids des mots. Le discours de gauche et de droite dans l'entre-deux-guerres. Maurice Thorez, Léon Blum, Pierre-Etienne Flandin et André Tardieu (1928-1939)*, Paris, 2000, Honoré Champion.

## The scientific values of lexicometry

To return to our point, we have tried to show in our doctoral thesis that lexicometry was of interest for the historian on two counts : one as proof and the other as a heuristic tool.

As a proof or aid to settle a number of points, let us take only two examples, quoting first an analysis by the great historian Louis Bodin on the French Communist Party's role in the Popular Front:

*It is remarkable that all the great mass demonstrations – a ritual to which the communists contributed unreservedly – took place before the outbreak of the Spanish war : February 12 1934, July 14 1935, February 16 1936 (...) They all aimed at uniting the people against the enemy within. The oath taken on July 14 1935 did not explicitly refer to external fascism. Obviously, the shift towards external antifascism did not take place, and it is debatable whether the Party really put all its forces behind it : The 200 families or Colonel de La Rocque were still more paramount in the Party's rhetoric than Hitler or Mussolini (...) The communists – in that respect not doing any better than the other left-wing forces – do not seem to have had a complete vision of the international situation, nor did they seem ready to march to the borders in defence of liberty.”<sup>4</sup>*

This excerpt firstly demonstrates the necessity for researchers in the fields of the Arts and History to rely on quantifiable data to draw irrefutable objective conclusions : at the heart of L. Bodin's argument, to establish the fact that internal antifascism superseded external antifascism, there is this assumption of a higher occurrence of “de La Rocque” over “Hitler” in communist discourse. Above all, the excerpt shows how, lacking any scientific means to quantify this value, one of the best specialists on the Popular Front and the role of the P.C.F., trusting only his own impressions, reaches a conclusion which goes against the truth on this essential historical point. No, in all the speeches of Maurice Thorez<sup>5</sup>, General Secretary of the P.C.F., the “200 families” (21 occurrences in the 1930s) are not quoted more often than “Mussolini” (106 occurrences); no, “de La Rocque” (29 occurrences) is not more present than “Hitler” (338 occurrences). We might add that French fascism, ten times less often mentioned in communist discourse than international fascism, is in most cases traced back to its foreign roots, de La Rocque, Taittinger and Maurras being repeatedly called “*agents of Hitler and Mussolini*”<sup>6</sup>. Thus Bodin's conclusions must be radically reversed. Yes, the P.C.F. did try to turn the internal antifascism born on February 6 1934 into an antifascist weapon to fight Hitler. In fact, for Thorez, it became very quickly the only meaning of the Popular Front, replacing its social meaning. And the slogan “Front Français” in the summer of 1936, which finds a natural expression in the communist wish to take part in the national union government of Thorez-Blum-Reynaud on the eve of Anschluss, shows unequivocally how the new party line of the P.C.F. was above all patriotic. With respect to the workers' movement the position of the Party between 1936 and 1938 can be criticised, but it is logical. Bodin, while pointing out that the P.C.F. gave up its social role and betrayed the revolution, suppresses the patriotic element, and thus leaves the baffled reader without an explanation.

<sup>4</sup> L. Bodin, “Le P.C.F. dans le Front populaire” in *Esprit*, 1966, p. 445-446.

<sup>5</sup> The Thorez corpus under study includes all the speeches and articles by Thorez between 1930 and 1939. We should remind the reader that in a party organised along ‘democratic centralism’, these speeches can be considered as the official voice of the Communist Party.

<sup>6</sup> M. Thorez, Speech given in the Wagram Hall, on December 26 1935. (Complete works, book II, vol. 10, p. 154-184).

Similarly Ilan Greilsammer, who insists on drawing a portrait of Blum as a charismatic leader asserts “he has the tone of conviction : he knows, is convinced, has the proof”<sup>7</sup>. The computer has no preconceived idea and studies all the verbs, all the words used by Blum indifferently, and shows precisely the opposite. If compared with two right-wing politicians like Flandin and Tardieu, or with an extreme left-winger like Thorez, Blum’s discourse is exactly that of the hypothetical, that of doubt, in which the “I think”, “I believe” and “I hope” appear in very significant statistical proportions. Hence, rather than as a charismatic leader it is possible to describe Blum as an intellectual lost in politics, whose hesitant and subtle dialectics were ill-adapted to his role as decision maker in the Manicheanism of a difficult historical period

The heuristic value of lexicometry justifies our interest even more. A computer, after having processed all the words, highlights a few for their quantitative characteristics; and then one has to find a historical explanation to objective lexical data that are sometimes unexpected.

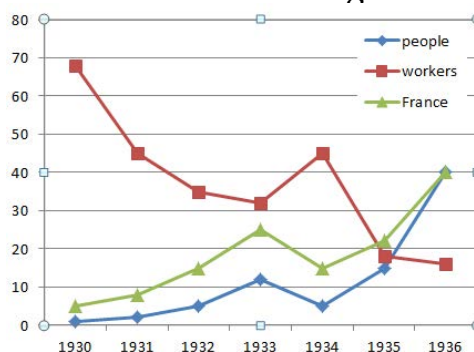
Why are “I” and “myself” statistically over-employed by Blum, just as they are by Jaurès and Mitterrand<sup>8</sup> ? Is it because an over-personalisation of politics and the concept of the exceptional individual are in reformist thought the only way to resolve the contradiction between legalistic republican practice and revolutionary marxist theory?

Why does the Right (Flandin and Tardieu) over-use “have” and “had” in the interwar period? Is it the trade mark of those who possess, or more simply the mark of a backward-looking conservative discourse, which multiplies the occurrences of the auxiliary verb for present and past perfect?

Why does a given word gradually disappear over the years from communist discourse after having been omnipresent ? ... So many questions thus present themselves faced with a hierarchical index of lexical items or a factorial analysis of concordances to stimulate historical reflection.

It is within this heuristic dimension that lexicometry has the most ambition : it claims to go beyond the hypothetico-deductive approach that has prevailed up to now in the analysis of historical texts. Where the historian was used to questioning the text, it is now the text which questions the historian. When a large corpus had to be dealt with, it was read with precise questions in mind, an established hypothesis, without which one became lost in the reading And yet the dangers of such a priori questioning are well known, as they often lead to certain answers while obfuscating others, which might be more relevant. Lexicometry dissects the text and produces a list of objectively discriminating terms which engender questions even before the historian’s subjectivity comes into play.

One example worthy of development concerns the history of the French Communist Party represented by a graph showing the frequency with which Thorez used the items “France”, “people” and “worker” during the thirties (January 1 1930 – December 31 1936)



<sup>7</sup> I. Greilsammer, *Blum, op. cit.*, p. 261.

<sup>8</sup> On Jaurès cf. P. Muller, *Jaurès, Vocabulaire et Rhétorique*, Paris, Klincksieck, 1994, p. 109 ff.; on Mitterrand cf. D. Labbé, *Le Vocabulaire de F. Mitterrand*, Paris, FNSP, 1990, p. 59 ff.

The computer measures the great turning point in the political stance of the communist party. There was a decline of the number of marxist words : “worker(s)” was used 6 times less in 1936 than in 1930 (12 occurrences per 10,000 words versus 68 per 10,000), and the same can be noted for “proletarian”, “proletariat”, “capitalism”, “struggle”, “bourgeoisie”, etc...Against this decline there was a considerable growth of patriotic and populist (in the strict sense of the word) vocabulary : “people” as seen on the diagram or also “France”, “nation”, “popular”, “democrats”, “parliament”, etc...

The computer also questions the historian on the timing of this evolution and sheds a new light on the history of the P.C.F. Was the lexical about-turn sudden and late in coming (1935-1936) as has been asserted<sup>9</sup> or gradual and as early as the beginning of the decade? Beside the two words illustrated here, we have been able to show that the lexical shift which abandoned the Bolshevic phraseology and led to the Frontist one began as early as 1930<sup>10</sup>.

The historian is also challenged by the computer on the question of 1934. There was an interruption of the trend, a break in the change, in 1934. Why?

This is not the time nor the place to answer the query, but the break has in effect hidden from the historian just how early and far-reaching the movement was, and has exaggerated the importance of the “big turn” of 1934-1936. In the end the question asked is deeper : was the Popular Front – born of a communist initiative uniting left-wing forces on a republican and patriotic platform – hatched in 1934, 1935 or previously in 1933 (if not even before) ? The latter supposition, in any case, takes into account some major progress in political lexicon which would have enabled the Popular Front’s stance to be put into words.

### **The resistance of the historical institution**

A tool for administering proof in certain debates or raising new questions : the method using lexicometry seems able to complete and strengthen our traditional reading. What then can we reproach it with, if only the introduction of a little rigour into a field which scarcely had any ? Three kinds of criticism have been levelled at it in the last twenty years, and their succession in time bears witness to the relentlessness with which the historical institution has repeatedly rejected it. Firstly, its validity was called into question : a purely lexical approach to discourse analysis was said to be insufficient, and the quantitative approach to the lexicon to be nonsense, in that semantics, the essence of language, lies in its quality, and particularly in its syntax. All lexicometrical studies have very early shown that the vocabulary played, if not a unique, at least a major role in political messages and that repetition was the prerequisite for the kind of linguistic effectiveness that a tract, an article, an electoral speech sought to achieve. What is more, today’s software programmes are no longer content with recording the repetition of vocabulary. The initial paradigmatic approach has now been combined with a syntagmatic lexicometry which systematically studies the words in their contexts : listing co-occurrences and ‘repeated segments’ and finally controlled consultation of phrases and paragraphs.

Then the readability of lexicometry was called into question and the assumed complexity of the method was put forward and presented as inaccessible to historians supposed to be impervious to statistics, IT and, why not, linguistics. Today’s lexicometric

---

<sup>9</sup> D. Peschanski, *And yet they do go round : the vocabulary and the strategy of the PCF (1934-1936)*, Paris, Inst. National de la langue française, Klincksieck, 1989.

<sup>10</sup> For the factorial analysis of concordances, cf. our thesis, vol. 2, p. 472.

processors are easier to use than word processors and the results, when structured around one or two statistical criteria, are simple to analyse. From these quantitative data the researcher can produce a historical analysis which should not hurt any 'traditional' reader.

The most recent criticism is based on the 'fruitlessness' of the method and more precisely (let us salute here the subtlety of the slip) of the ratio of pluridisciplinary investment versus results. This brings us to the heartfelt confession of the most eminent member of our jury at our *viva voce* presentation. "I measure the merits of a method first and foremost by its fruitfulness", he asserted, hinting perhaps that the lexicometric tree whose roots and trunk were allowed to remain was not bearing enough fruit.

With all due respect, we seriously disagree with this assertion for it implies that the historian can be satisfied with a fruitful, fertile and prolific method...which would only give birth to monsters. In science the most important criterion is that of the validity of the results. In History productivism is obviously necessary but always secondary, unless we, once again, complacently renounce the status of "science" and embrace the journalistic approach in which print runs and sales are more important than the contents exposed. The historian must measure his method primarily by the validity of his results.

Above all, the fruitfulness of lexicometry need no longer be demonstrated for even when dealing with a body of well-known and often analysed speeches, such as the PCF speeches, it is still possible today, as illustrated above, to provide new insights, and rectify chronologies. For our own part, we have also studied a corpus of entirely unpublished right-wing speeches, for which the lexicometric method proved to be reliable, rapid and stimulating in marking out virgin territory. Later on, when dealing with the new piles of documentation in macro-corpora made instantly available – e.g. the whole of the Fifth Republic Journal Officiel – it is a safe bet that only the computer will enable us to explore horizons that the human eye cannot embrace. The ratio, – since it is seen in these terms – between methodological investment and historical results might very well be reversed between traditional and computer aided reading. And let us beware of banning the former !

\*  
\*       \*  
\*

This plea in favour of lexicometry does not aim at making converts. France is not yet ready for the dissemination of a method which remains limited in History<sup>11</sup>. We simply wished to demand the right for more rigour where subjectivity is being imposed. On the other hand, the militant refusal, these days, of scientific, methodical approaches in History raises a fundamental question about current historiography, which sets the relativity of the researcher, and thus relativity of historical truth itself, as the ultimate end.

Actively rejecting scientific methods, recoiling complacently into subjectivity, advocating the intuitive approach not as a necessary stopgap but as a real panacea, is forbidding oneself to bear judgement on things, men and History; it is forbidding oneself to escape from preconceptions, that is to say the dominant ideology which constitutes the basis for our working hypotheses and our historical conclusions. If we keep on repeating that there is no absolute truth, are not we condemning ourselves to remain for ever mistaken ? "Everything is relative", "there are as many truths as there

---

<sup>11</sup> Philippe Genet was noting during the 2<sup>nd</sup> Conference of the French Association for History and IT : "*Quantitative lexicology is a methodological tool the resources of which historians are not yet using... these methods have not yet found their way into the French historians' everyday tool-box.*" (in *Cahiers de la Méditerranée*, n° 53, December 1996, "History and IT", p. 123.

are researchers”, “for History no one is responsible, no one is guilty” : the policy of the present political history is clear. The historian, just like the citizen, whatever his personal mood, has neither the right nor the duty to resort to scientific tools to formulate judgement. Should he not also give up trying to understand, come to conclusions and, if need be, rise up in protest ?