

INTRODUCTION

SLIDE 1

The alliance of textual statistic methods and discourse analysis has long been used in the field of political and media discourse studies. This paper applies such methods to the analysis of different draft versions of social reports in attempt to uncover the pragmatic goals behind this specific type of discourse construction. Moreover, this paper is a continuation of the earlier Veniard study on lexicon used by social workers in their reports to describe “what is wrong” with the child and his/her family (Veniard 2008).

SLIDE 2 -HYPOTHESIS

To start with, we formulate a main work hypothesis on the textometric explorations of the writing process. Observation of repetition in the writing process reveals both discursive routines which highly mark the reports as well as local changes that affect the writing process. These two phenomena display the strategies that characterize the processing of each child-report.

We organized the textometric exploration around two points of entry:

- ⇒ the corpus as a whole unit: two types of phenomena are observed: (1) the frequency with which vocabulary appears and evolves throughout the corpus and (2) the structure of different sections of the reports themselves.
- ⇒ the second point of entry is a set of all-encompassing words that denote “what’s wrong” with the child and his family. For example: *problème(s)*, *des difficultés*, *trouble(s)*, *symptôme* (problem(s), some difficulties, symptome, disorder(s)).

From a textual point of view, these words are key-words weaving the social workers’ discourse. **From a syntagmatic point of view**, these items can be considered as construct nouns or predicative words that have an opening on utterance and discourse. **From a semantic point of view**, these words are polysemic and vague. In fact, people who work in (juvenile) justice criticize their use and advise social workers to give an accurate description of what is wrong. Therefore, we can investigate if the social worker make such a description when using this vocabulary and then evaluate how the usage of these words responds to the main pragmatic goals attached to these social reports, namely: submitting before a judge the dangers run by the child and accompanying the child and his/her family without upsetting/harming either.

Therefore, this starting point is productive not only for the analysis of the child’s and family’s situation as described by the writer but for the detection of different types of phenomena and strategies concerning the writing process too.

SLIDE 3

As a first step, we review textometric methods and explain how we applied them to a corpus of social report drafts. As a second step, we focus on discursive routines and strategies the textometric explorations have lead to reveal.

II- Textometric methods

Textometry stems from a relatively recent movement in statistics applied to the study of language. Literary statistical studies (Yule, 1944), correspondence analyses (Benzécri, J.P, 1973), and chronological textual analyses (Salem, 1991) are among some of the pioneers in this multidimensional field. This movement has its foundations in the statistical study of linguistic data, originally used for lexical counts of words. In this approach, the text is considered as possessing its own internal structure, comparable to a prism of numbers and figures, producing information on the frequency counts of words which allows for the analysis of larger amounts of text than could be analyzed by manual means alone.

II.A Textometric objects and process

First, depending on the question at hand, the corpus is broken down into different sub-sections that will allow the analyst to visualize important variations across these different zones. In the case of worker's drafts, the corpus is separated into the different versions.

SLIDE 4

Therefore, a different sub-section is created for each draft in chronological order. Tags are used to mark each zone for computer analysis in the textometric softwares Le Trameur¹ & Lexico 3².

SLIDE 5

[SLIDE 6 ARBORESCENCE]

Second, the process of transforming a text into an analyzable sequence of units by computer assisted methods is not trivial. In order to obtain units of count, the text must be segmented into the smallest components that are to be analyzed.

SLIDE 7

The segmentation plan is determined by the analyst and breaks up the text into occurrences (tokens), single graphical instance of a word and forms (types)³, a single graphical unit corresponding to several instances. For example, in the corpus of writer's drafts, the word

¹ Le Trameur developed by S. Fleury, SYLED Sorbonne Nouvelle <http://www.tal.univ-paris3.fr/trameur/>

² Lexico 3 developed by A. Salem, SYLED Sorbonne Nouvelle <http://www.tal.univ-paris3.fr/lexico/lexico3.htm>

³ Here we adopt the terminologie used by Lebart & Salem, 1994, Exploring Textual Data.

“difficulté” may be found across several versions and/or several paragraphs, these different instances are called occurrences; however, the frequency dictionary will only display one form “difficulté”.

SLIDE 8

Forms can then be studied through the distribution of the number of their occurrences across the various zones of the corpus, as can be seen in figure 1, for example.

SLIDE 9

Moreover, more complex probabilistic calculations can be done once the corpus has been segmented.

II.B Co-occurrence method

SLIDE 10

The co-occurrence is one such complex unit of count that corresponds to the statistical attraction of two or more words in a given span of text (sentence, paragraph, entire article report). This analysis allows for the precise description of the lexical environment of a pivot-form (or pivot-word). Though various definitions of co-occurrences exist in statistical analyses of text, here we use the co-occurrence method as described by Lafon (1980) and later Martinez (2000). The hypergeometric distribution is applied to calculate the lexical associations of a pivot-form, in which several variables are left to the end-user (Martinez, 2000). First, the co-frequency of two associations must be determined; this frequency indicates the lowest number of times two forms appear together in the corpus, in the defined context. Second, a threshold is provided, designating the probability level that co-occurrence relationship must have for appearing in the predefined context (Lebart & Salem, 1994).

What results is a list or network of co-occurring forms that can be interpreted through the following:

- Frequency: the total frequency of the co-occurrence in the corpus
- Co-Frequency: the frequency with which the co-occurrence appears with the pivot-type in the defined context
- Specificity: the degree of probability that the co-occurrence will appear in that context
- Number of contexts: the number of contexts that the co-occurrence and pivot-type appear together in.

This unit of analysis seems particularly interesting for detecting associative relationships between words.

SLIDE 11

In taking co-occurrence analysis one step further, it is also possible to calculate polyco-occurrences (Martinez, 2003), otherwise known as the co-occurrences of co-occurrences.

After calculating the network for a given pivot-form, each resulting co-occurrence is then analyzed itself as a pivot-form in the same context as the original pivot, producing a network of interrelated units. These interrelated units serve as a guide for observing unexpected or evolving textual relationships between a pivot and the resulting co-occurrences. Analysis requires greater context of the sequence or sequences containing them.

SLIDE 12

It is therefore necessary to observe the network of units in the original context either by using a concordance or map of sections, giving access to the sentence or paragraphs the units are found in.

II – Discursive Routines, local changes and common strategies of the social writing/writers

The “écrits au travail” or “writings at work” (Boutet, Franckael, Delcambre 1992) involve some discursive routines. This property can be observed in the social reports through detection of typical usages and sequences. The genetic and textometric analysis of social reports, by comparing different draft versions for each report, reveals the routine aspect of this discourse from its writing process.

III.A. Repetition & stability

First, we can observe that one part of the organization of the text remains stable for each report and for each draft version.

SLIDE 13

A large group of sections appears for each report, and we do not notice any change for these sections from one draft version to another. Furthermore, a few sections have similar names.

Secondly, distribution of high frequencies and low frequencies indicates for each report very little variations and diversification of vocabulary from one draft to another and from one section to another. In other words, the texts in the reports are very repetitive. Social workers begin their writings with a pre-established structure and use words and expressions from a common lexicon. In a way, the text has the face of “jargons”: languages with their own codes and constraints.

3.2. Repetition & phraseology

Observation of the frequencies and co-occurrences of the words describing what is wrong leads to detect a phraseology of the social worker.

- **repetition / frequencies of key-words that get denser => “phraseological effect ”**

For the noun category, repetition even creates a « phraseological effect » (Fiala 1987: 36). And we can see a densification of nouns that indicate in the reports “what is wrong”.

SLIDE 14

This phenomenon is very striking especially with the forms *difficulties* and *difficulty* in both final reports. The densification of the word *difficulty* comes with the progressive use of derivatives in the text: *difficile* (*difficult*) and *difficilement* (*with difficulty*).

SLIDE 15 EXEMPLES

In several examples of the corpus, we can even see that, after several drafts, when the noun *difficulty* is used, the author starts using its morphological derivatives at the same time.

I will not read the exemple but you can see the occurrences of derivatives of *difficulty* are frequent in the corpus.

« Dans les faits elle capte très **difficilement** l'attention de son père très envahit par ses propres **difficultés**. » (Rapport Thierry Roux, état 5/16)

« Thierry a beaucoup grandi durant cette année, il n'est pas toujours très à l'aise avec sa taille et se tient **difficilement** droit. Son allure générale laissait penser à des **difficultés** au niveau du dos ». (Rapport Thierry Roux Etat 5/16)

« Scolarité La scolarité est toujours aussi **difficile**. La question de la déficience repérée par notre psychologue n'a pas été traitée avant octobre. [...] la rencontre avec le directeur et la surveillante principale a montré que les enseignants prenaient actes des **difficultés** d'Annie mais ils n'avaient pas pu la mettre en classe allégé en raison de son choix de langue puis pour l'entraîner vers un niveau supérieur. (Rapport Annie Pauty-Etat 9/12)

▪ phraseological units/items : **fixed constructions et discursive circularity**

Research from Saint-Cloud laboratory defined phraseological units as recurring combination of lexical & grammatical forms, more or less stabilized. Observation of co-occurrences around the word *difficulty* helps detect the recurrence and the stability of some constructions in the writing process.

SLIDE 16

For instance, we can notice that this word is quickly associated with the same prepositions (“en” *in/with*), « dans » (*in/with*), « pour » (for) as well as recurrent relational adjectives that can be “*academic*”, “*maternal*”. Nevertheless, as far as frequencies are concerned, these associations with prepositions remain relatively stable from one draft version to another.

SLIDE 17

The contexts in which the word *difficulty* is used show that the associations with prepositions and relational adjectives refer to constructions and lexical uses more or less fixed. We can find these constructions in all reports we have studied, and some of them can be compared to constructions in language dictionaries. **CLIC** For exemple: “*avoir une/quelque difficulté*

à/pour faire qqch” or “to have (a) difficulty to do something”. However, most of them seem to be phraseological units that are specific to social work jargon.

When analyzing cooccurrences, we can also detect a more local phraseology, specific to each report considered separately.

SLIDE 18

First we observe recurring associations that correspond either to the repetition of the same constructions by the writer or to the re-using of the same sequence with little changes.

For instance, in Thierry Roux's report, we can see a very specific phrase “the father is overcome by his own difficulties” over three draft versions. This phrase forms a kind of a single block which is re-used to describe all the relationships between the children and their parents.

Secondly, we can note unusual constructions. In Thierry Roux’s report, the social worker uses the expression “**difficulties of the back**” instead of the everyday expression “**backache**”. This unusual use can certainly be analysed as an effort (from the social worker) to give to the judge an objective and processive evaluation. But we can interpret it as the result of an automatism as well. In fact, this use gives the impression that the social worker cannot help using the word *difficulty* in his evaluation. These local phenomena display for each report an automatic and mimetic writing.

3.3. Repetition & insistance

As we were analysing frequencies and co-occurrences around the word *difficulty*, we have been conducted to go beyond the problematics of phraseology. In fact, we detected a strategy built by social workers in their description of what is wrong.

We wondered about the repeated and « invading » presence of the word *difficulté* in reports that are supposed to be very accurate. Is that word systematically linked to elements of specification?

SLIDE 19

At the same time, the graphs of co-cooccurrences alerted us to an interesting phenomenon in Thierry Roux’s report. The pivot-form *difficulties* has a preferred co-occurent in the sentences of the report, the intensive adverb *very*. This strong relationship between *difficulties* and *very* occurs from the first steps of the writing process. The graphs of co-occurrences around the pivot-form *very* confirm the reciprocity of this “attraction”. Moreover, unlike the associative relationships between the pivot-form and prepositions, this relationship becomes stronger as the writing progresses. The frequency of this co-occurrence, and the number of contexts in which the co-occurrence appears, even indicate that the adverb is repeated in some

contexts with the word *difficulty*. Indeed, a return to the text is instructive to understand what is the stake with the use of the word *difficulties* in final reports and, more generally, with the use of words that denote what is wrong.

SLIDE 20

First, we can observe that the word *difficulté* is particularly specified when it referentially applies to the child. From the syntactic point of view, the elements of specification appear as prepositional complements, relative adjectives or independent sentences. Nevertheless, we can notice these elements of specification are situated in the right co-text of the word *difficulté*, after the naming of what is wrong.

Furthermore, when these elements are used in the same sentences as *difficulté*, they often appear in a right-detached position:

En troisième, Evelyne a rencontré **des difficultés** croissantes, dans les apprentissages, notamment **en mathématiques** et **en anglais**.

When the writer names parents difficulties, he omits nearly systematically to give an accurate description of the problems encountered (2 specifications for 9 occurrences, and one of the specification in parenthesis). Meanwhile, at this moment of his writing, the writer always uses markers of intensity: adverbs *very*, *extremely*, adjectives *real*, *important*: [CLIC]

Madame est en grande difficulté pour se mobiliser autour des besoins de ses enfants et de l'exercice de son autorité parentale.

In some sentences, the writer uses both markers of intensity and markers of insistence such as *above all*, *mainly*: [CLIC]

En effet, **Monsieur** est **très envahi** par ses **difficultés propres** et s'il revendique les droits de visite, il les investit **surtout** comme un lieu de parole pour lui.

In this way, the social worker seems to underline a problem that he doesn't name clearly. This phenomenon of insistence on the naming of parents difficulties covers the report as far as its conclusion which recommends insistently to maintain the placement of the children:

[SLIDE 21]

Les relations des deux parents avec leurs enfants sont **très** anxiogènes pour eux, chacun à leur niveau **en fonction de ses difficultés propres** il ressort des deux années écoulées une évolution des quatre enfants, qui reste **extrêmement** fragile, empreinte d'un fonctionnement familial qui remet en cause régulièrement leur développement psychique et nécessite un étayage éducatif et psychologique **très important**. Il est fondamental de maintenir le dispositif actuel de placement qui garantit la protection des enfants.

We can observe a phenomenon of this sort in Marine Dumont's final report: in fact, from the early draft versions, the word *difficultés* is associated with the adverb *toujours* (*still*).

Finally, this local phenomenon of insistence leads to reconsider the phenomena of repetition and cycle-like behaviour we have described before. We can't put discourse cycle-like behaviour, repetition and densification of the word *difficulté* down only to a professional phraseology. These phenomena can also be observed in the progressive construction by the social worker of an alert system. In a sensitive situation of communication, this process combines both partial description of what is wrong and insistence on what is wrong.

SLIDE 22 - CONCLUSION

Methodology:

- Textometry sheds light on the inherent cycle-like behavior of discourse allowing us to dynamically reveal strategies to get one's message across or convince a judge of the best possible future for the child being discussed.
- Statistics make visible these discursive patterns that could have otherwise gone undetected if we were simply to read the reports. This is the case with the "insistence" phenomena seen in the presentation

From a pragmatic point of view three important conclusions that require further investigation:

- The social workers leave discursive "clues" to the judge while trying to code them in such a way that they would not be hurtful to the families or child they concern. This is visible in the use of the repeated word "difficult" which is quite vague. The repetition is way of applying more weight to the reasons associated with this word. Moreover, an interview of the social worker's perception of their writing called words such as "difficuty" or "problem", "darwer-words" a sort of hyperonym in which they can refer to a variety of non descript delicate situations.
- But, these reports are also part of the "job description", meaning they are a routine part of any social worker's daily tasks. This very nature of writing reports may generate a certain number of recurrent codes in the writing process. These codes do not always correspond to what the "receiver" in this case judge, is expecting, namely an accurate description and evaluation of the situation.
- Finally, we also think the use of the such vague fixed expression, beyond being part of the professional jargon, could be a way for the social workers to emotionally distance themselves from the situation they are writing about.

SLIDE 23

Bibliographie

Branca-Rosoff, S., Torre, V., (1994), « Observer et aider : l'écrit des assistantes sociales dans les 'Demandes d'interventions' », *RSFP* n°12, Université de Provence, p. 115-135.

Fleury, S. (2007), « *Le Trameur, Manuel d'utilisation* » (mise à jour 2009), disponible sur <http://tal.univ-paris3.fr/trameur/leMetierLexicométrique.pdf>

Branca-Rosoff, S. (coord.), (1999), *Langage et société*, n°87, mars 1999, « Types, modes et genres de discours ».

Lamalle, C., Salem, A. (2002), « Types généralisés et topographie textuelle dans l'analyse quantitative des corpus textuels » [en ligne], Inria, St Malo, in *Actes des 6e journées d'analyse statistique des données textuelles*, disponible sur internet : <http://www.cavi.univparis3.fr/lexicométrica/jadt/jadt2002/tocJADT2002.htm>

Lamalle, C., Martinez W., Fleury, S., Salem, A., Fracchiola, B., Kuncova, A., Maisondieu, A. (2003), *Lexico 3. Outil de statistique textuelle. Manuel d'utilisation*, Paris, SYLEDCLA2T-université de la Sorbonne Nouvelle-Paris 3, 50 p, disponible sur : <http://www.cavi.univ-paris3.fr/Ilpga/ilpga/tal/lexicoWWW/manuel3.htm>.

Lebart, L., Salem, A. (1994). *Statistique Textuelle*, Paris, Dunod.

MacMurray, E. Shen, L. (2010), "Textual Statistics and Information Discovery: Using Co-occurrences to Detect Events" in *Actes du Colloque VSST'2010, Toulouse*.

Martinez W., (2003) *Contribution à une méthodologie de l'analyse des cooccurrences lexicales multiples dans les corpus textuels*. Doctoral Dissertation in Language Science, Université de la Sorbonne nouvelle - Paris 3.

Mellet, S., Salem, A. (éds.), (2009), *Lexicométrica* n° spécial « [Topographie et topologie textuelles](#) », nov. 2009, disponible sur : <http://www.cavi.univ-paris3.fr/lexicométrica/numspeciaux/special9.htm>.

Née, E., (2009), « *Insécurité* et élections présidentielles dans le journal *Le Monde* », *Lexicométrica*, numéro thématique « Explorations textométriques », volume 1 : « Corpus et problèmes », S. Fleury, A. Salem (dir.), 2009, disponible sur : <http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicoWWW/navigation/Volume1.pdf>.

Née, E. (2007), « Les outils lexicométriques à l'épreuve d'un corpus médiatique », *Les corpus en linguistique et en traductologie*, M. Ballard, C. Pineira (Éds), Arras, Artois Presses Université, 2007, p.71-88.

Pétillon, S., Garnier, F. (éds.), 2006, *Langages* 164 : « La révision de texte : méthodes, outils et processus », Paris, Larousse/Armand Colin.

Salem, A. (1988), "Approches du temps lexical. Statistique textuelle et séries chronologiques", *Mots* N°17, octobre 1988, p. 105-143.

Salem, A., Fleury, S. (2009), *Lexicométrica* n° Spécial « [Explorations textométriques](#) », [Volume 1 : Corpus et problèmes](#), [Volume 2 : Séries textuelles chronologiques](#), [Volume 3 : Corpus multilingues](#), nov. 2009.

Veniard, M., (2008), « Écrire "ce qui ne va pas" dans le champ de l'enfance en danger : les mots *problème(s)* et *difficulté(s)* », *Les carnets du Cediscor* 10, Presses Sorbonne Nouvelle.