

Textometric Explorations of Writing Processes: A Discursive and Genetic Approach to the Study of Drafts

Emilie Née¹, Erin MacMurray², Serge Fleury³

¹Université Paris 3 Sorbonne Nouvelle – emilie.nee@gmail.com

²Université Paris 3 Sorbonne Nouvelle – erin.macmurray@gmail.com

³Université Paris 3 Sorbonne Nouvelle – serge.fleury@univ-paris3.fr

Abstract

The Textometric method, the observation of variations, has yet to be applied to the study of the writing process, the different stages of writing the same text. This paper uses Textometry for the study of four social workers' reports on children "at risk". The Textometric method is used to analyze the discursive and pragmatic patterns employed by the social workers and their description of "what's wrong" with the child at various stages of the writing process.

Résumé

La méthode Textométrique n'a pas été appliquée encore à l'étude des processus d'écriture, des différentes versions d'un même texte. Cette étude utilise la Textométrie pour l'analyse de quatre rapports des travailleurs sociaux sur des enfants « en danger ». La méthode Textométrique est utilisée dans l'analyse des stratégies discursives et pragmatiques mises en place par les travailleurs sociaux lors de leur description de « ce qui ne va pas » avec l'enfant au cours de différents états du processus d'écriture.

Keywords: Textometry, co-occurrences, writing process, discourse analysis

1. Introduction

The alliance of textual statistic methods and discourse analysis has long been used in the field of political, media, and more recently scientific discourse studies. No current Textometric methodology exists to study writing processes, meaning the discursive and genetic analysis of producing different versions of the same text. The first objective of this paper is to apply Textometric methods to the analysis of different draft versions of social reports¹ in attempt to uncover the pragmatic goals behind this specific type of discourse construction. In this objective the pragmatic challenges inherent to this specific type of text must be taken into account: 1) submitting before a judge the risks/dangers run by the child; 2) accompanying the child and his/her family without hurting either; 3) avoiding confining the child in a specific type of social, judicial or medical discourse. Moreover, this paper is a continuation of the earlier study on lexicon used by social workers in their reports to describe "what is wrong"

¹ This paper was prepared within the framework of the project ANR ECRITURES « Brouillons d'écrits sociaux : approche génétique, discursive et textométrique de l'écriture professionnelle » (Drafts of Social Workers' Writings: a Genetic, Discursive and Textometric Study of Professional Writing, <http://www.univ-paris3.fr/anr-ecritures/>)

with the child and his/her family (Veniard, 2008). The second objective of this study is to formulate a Textometric methodology for the study of the writing process.

Observation of repetition in the writing process may reveal both macro discursive routines that highly mark the overall structure of the reports as well as local changes from draft to draft that affect the writing process. These two phenomena display strategies that characterize the processing of each child-report and we hypothesize that they can be quantifiably observed through Textometric methods.

2. Applying Textometry to the study of drafts

Four social reports were used as the corpus for this initial explorative study of writing processes. The names of the children and their families were changed in all reports in order to maintain anonymity. Two of the reports correspond to Synthesis Reports (*Documents de Synthèse*) that review the child's progress over the a given period of time and the other two reports correspond to Final Reports (*Rapports d'échéance*) that summarize the child's evolution². The corpora Annie-Pauty and Anthony-Viti are examples of Synthesis Reports whereas Thierry-Roux and Marine-Dumont are examples of Final Reports. Each corpus was separated into the different draft versions from the first rough draft to the final draft of the social worker's report as in the figure 1 below. The draft versions were collected by the social workers themselves. Each draft corresponds to a version saved electronically on their computer. The version of the document was indicated using the tag `<etat= n >` for processing in Textometric software *Le Trameur* (Fleury, 2007) as well as *Lexico 3* (Salem et al, 2003).

Prior to the textometric analysis, the drafts of the corpus were aligned (figure 1) in order to visualize changes in the writing process³. This helps target blocks of text that are deleted or added as the drafts progress.

The Textometric exploration was organized around two points of entry:

- the corpus as a whole unit: two types of phenomena are observed: (1) the structure of different sections of the reports themselves and (2) the frequency with which vocabulary appears and evolves throughout the corpus;
- 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, (juvenile) justice norms criticize their use and advise social workers to give an accurate description of what is wrong (Huyette, 2003). Therefore, such descriptions made by the social worker when using this vocabulary

² Each report is written by one or more social workers. The final draft is proof read by the head of the department.

³ The drafts were aligned for textometric exploration in *mksalign* (Fleury & Zimina, 2005) allowing the comparison of draft pairs. The complete alignment is displayed on the ANR Ecritures website: http://www.univ-paris3.fr/66256407/0/fiche_pagelibre/&RH=1295620557102

can be investigated and evaluated as to how the usage of these words respond to the main pragmatic goals attached to these social reports mentioned above.

This point of entry is therefore productive not only for the analysis of the child's and family's situation as described by the social worker writing the report, but also for the detection of the different types of phenomena and strategies concerning the writing process.

	Draft 1	Draft 2	Draft 3	Draft 4	Draft 5	Draft 6	Draft 7	Draft 8	Draft 9	Draft 10
14	Marine a des difficultés en espagnol, matière qu'elle vient de commencer cette année mais de façon générale elle se montre volontaire même si son assistante familiale doit parfois lui rappeler ses obligations scolaires.	Marine a des difficultés en espagnol, matière qu'elle vient de commencer cette année mais de façon générale elle se montre volontaire même si son assistante familiale doit parfois lui rappeler ses obligations scolaires.	Marine a des difficultés en espagnol, matière qu'elle vient de commencer cette année mais de façon générale elle se montre volontaire même si son assistante familiale doit parfois lui rappeler ses obligations scolaires.	Marine a des difficultés en espagnol, matière qu'elle vient de commencer cette année mais de façon générale elle se montre volontaire même si son assistante familiale doit parfois lui rappeler ses obligations scolaires.	Marine a des difficultés en espagnol, matière qu'elle vient de commencer cette année mais de façon générale elle se montre volontaire même si son assistante familiale doit parfois lui rappeler ses obligations scolaires.	Marine a des difficultés en espagnol, matière qu'elle vient de commencer cette année mais de façon générale elle se montre volontaire même si son assistante familiale doit parfois lui rappeler ses obligations scolaires.	Marine a des difficultés en espagnol, matière qu'elle vient de commencer cette année mais de façon générale elle se montre volontaire même si son assistante familiale doit parfois lui rappeler ses obligations scolaires.	Marine éprouve des difficultés en espagnol, matière qu'elle vient d'appréhender cette année.	Marine éprouve des difficultés en espagnol, matière qu'elle vient d'appréhender cette année.	Marine éprouve des difficultés en espagnol, matière qu'elle vient d'appréhender cette année.
	Different stages of writings									

Figure 1 – Example of different versions of the report Marine-Dumont

The reports had a total of 62 drafts for 114,724 tokens and 4036 types. Their specific Textometric characteristics are summarized in table 1 below.

Table 1- Textometric characteristics of the social workers' reports

Corpus	Drafts	Tokens	Types	Type Max
Thierry-Roux	16	71101	1860	de (Preposition)
Marine-Dumont	10	14045	848	de (Preposition)
Annie-Pauty	12	8846	638	de (Preposition)
Anthony-Viti	24	20732	690	de (Preposition)

Several Textometric methods were used to explore the discursive routines in the different reports. 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 report). This analysis allows for the precise description of the lexical environment of a pivot-form (or pivot-type). 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 more recently Martinez (2003). The hypergeometric distribution is applied to calculate the lexical associations of a pivot-form, in which several variables are left to the analyst (Martinez, 2003). 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 parameters:

- **Frequency**: the total frequency of the type in the corpus;
- **Co-Frequency**: the frequency with which the pivot-type and the co-occurrence simultaneously appear in the defined context;
- **Specificity**: index indicating the over-representation (or under-representation) of the co-occurrence in the defined context in relationship to the pivot-type and the co-occurrence in separate contexts;
- **Number of contexts**: the number of contexts in which the pivot-type and the co-occurrence appear simultaneously.

This unit of analysis seems particularly interesting for detecting associative relationships between words. 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. 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 where the co-occurring units are found.

3. Textometric Exploration of Discursive Routines

Discursive Routines (Boutet et al., 1992) 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.

3.1. Repetition and Stability

First, the logical organization of the text can be observed (Lebfevre, 2011). The text is often *structured* by the social worker into various sections in the first draft; these sections are then filled during the writing process. A certain number of these sections remain stable for each report and for each draft version. Furthermore, the sections are often indicated in the same way from report to another. The sections given in table 2 for the Synthesis Reports account for half of the sections that are used in the Annie-Pauty corpus and for six sections in the Anthony-Viti corpus. Concerning the Final reports, all the sections noted account for one third of the sections in the Thierry-Roux corpus as well as half the sections in the corpus for Marine-Dumont, thus showing the stability of these pre-defined categories used to organize the report. The sections that were modified often pertained to background information specific to the child in the report. Such sections were removed from final drafts in order to maintain only neutral information that would not hurt the child who may access the report later in life.

Table 2- Spatial organization of the drafts (Sections that are common to all four reports)

Synthesis Report (<i>Document de Synthèse</i>)	Final Report (<i>Rapport d'échéance</i>)
Progress (<i>évolution</i>), Care (<i>prise en charge</i>), Environment (<i>environnement</i>), Health (<i>santé</i>), Education (<i>scolarité</i>)	Medical Care (<i>soins</i>), Education (<i>scolarité</i>), Child + NP (<i>Enfant + NP</i>), Progress (<i>évolution</i>), Conclusion (<i>conclusion</i>)

The progression of the different sections is shown for the Annie-Pauty report in figure 2. Between draft version 1 and draft version 12 it is clear that the sections put in place by the social worker are completely modified. By draft version 8 all the sections in the final draft are present and remain stable until the end. The different levels shown in this figure represent the logical organization established by the social worker of the different sections, much like the structure of a reported document⁴. This spatial organization of the report was observed in all four cases.

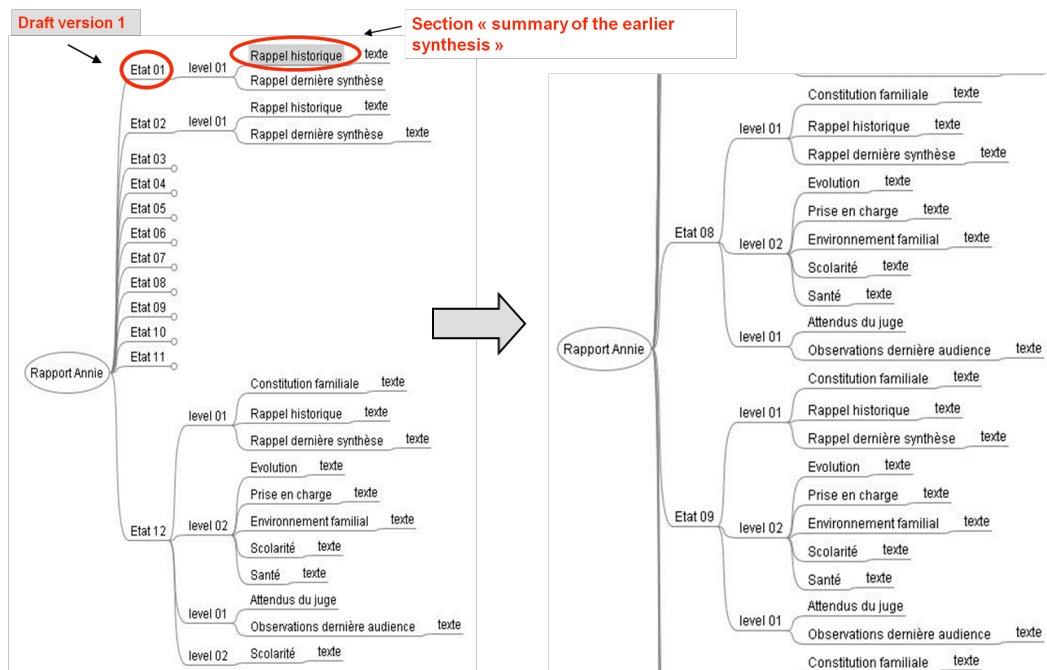


Figure 2 – Tree Diagram of Sections for Annie-Pauty

Secondly, the distribution of frequencies in the vocabulary increase figure 3 below indicates very little variation and diversification of vocabulary from one draft to another and from one section to another. This phenomenon is common to all four reports observed. 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. This leads to the hypothesis that the text is composed of “jargons”: languages with their own codes and constraints.

⁴ Levels 1 and 2 are displayed as sister nodes in the tree structure (figure 2). This reflects the tag organization used for textometric analysis and not the actual imbrications of the hierarchical structure.

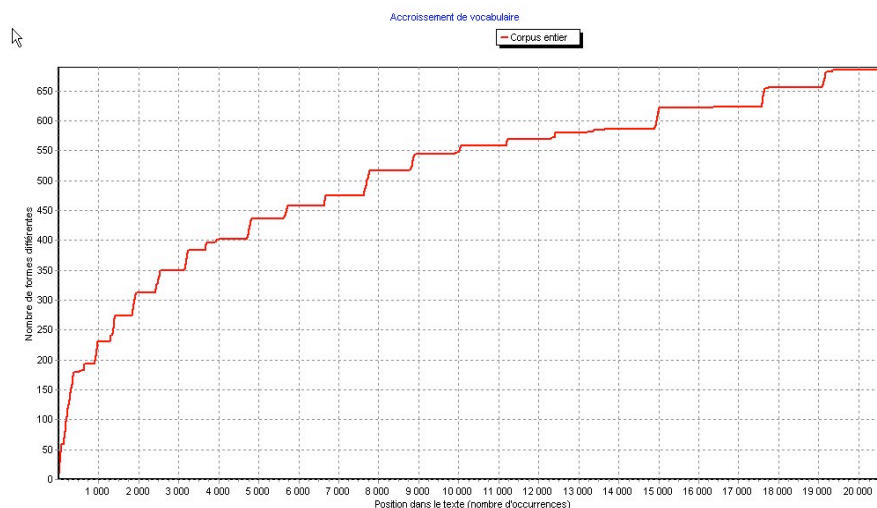


Figure 3 – Vocabulary Increase for Anthony-Viti Synthesis Report

3.2. Repetition and phraseology

Observation of the frequencies and co-occurrences of the words describing “what is wrong” shows the specific phraseology of the social worker being developed in the writing process. Specifically for the noun category, repetition creates what is called a “phraseological effect” (Fiala 1987: 36). A densification of nouns that indicate “what is wrong” can be observed in all four reports. This phenomenon is very striking especially with the forms *difficulties* and *difficulty* in both Final Reports, figure 4. The densification of the word *difficulty* comes with the progressive use of derivatives in the text: *difficile* (difficult) and *difficilement* (with difficulty).

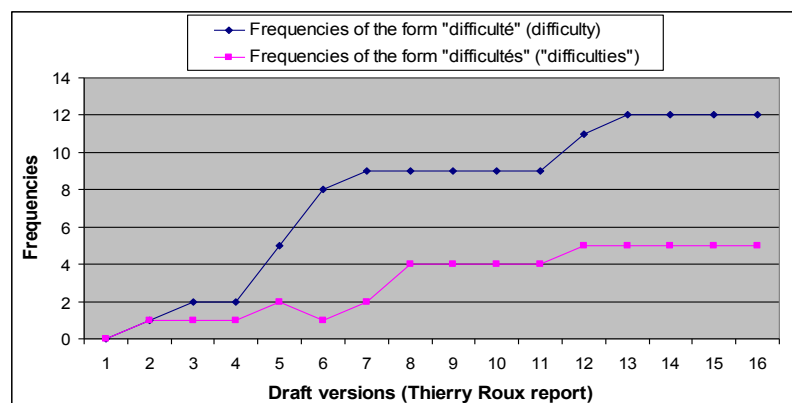


Figure 4- Frequency of forms *difficulté* (difficulty) and *difficultés* (difficulties) in Thierry-Roux

In several examples of the corpus over the course of several drafts, when the noun difficulty is used, the author progressively uses its morphological derivatives at the same time. Such occurrences of derivatives of *difficulty* are frequent in the corpora.

- i. [Thierry-Roux, drafts 5-16] Dans les faits elle capte très **difficilement** l'attention de son père très envahit par ses propres **difficultés**.

she catches « **with difficulties** » her father's attention who is overcome by his own **difficulties**

- ii. [Thierry-Roux, drafts 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.

Thierry has grown a lot this year, he does not feel very comfortable with his height and stands up straight "**with difficulty**" leading us to believe he has "**difficulties of the back**".

- iii. [Annie-Pauty, drafts 9-12] 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 acte des **difficultés** d'Annie mais ils n'avaient pas pu la mettre en classe allégée en raison de son choix de langue puis pour l'entraîner vers un niveau supérieur.

She still has many **difficulties** at school. [...] The meeting with the headmaster and the chief supervisor showed that the teachers noted Annie's **difficulties** [...]

Phraseological units can be defined as recurring combination of lexical and grammatical forms, more or less stabilized (Branca-Rosoff, 1997; Fiala et al., 1987; Lafon, 1997; Salem, 1987). Observation of co-occurrences of the pivot-type *difficulté* (*difficulty*) helps detect the recurrence and the stability of these constructions in the writing process, figure 4. This type is frequently 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.

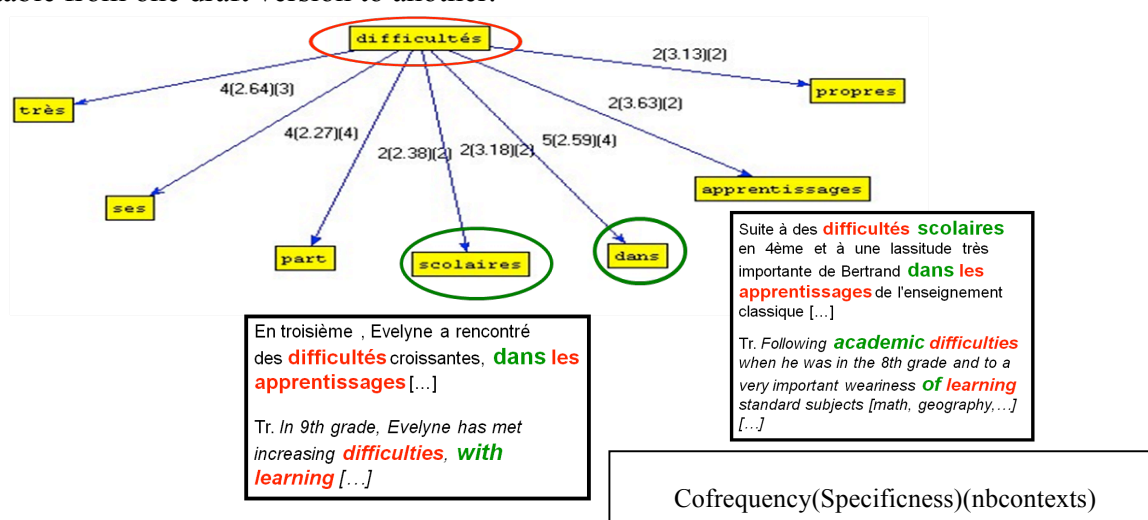


Figure 5- co-occurrences for the pivot-type *difficultés* (difficulties) and prepositions Thierry-Roux draft, parameters: co-frequency 2, threshold 2

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 (examples iv-vii). These constructions are found in all reports studied. They can be compared to constructions seen in language dictionaries, for example: "avoir une/quelque *difficulté* à/pour

faire qqch” or “to have (a) difficulty to do something”. However, most of these constructions appear to be phraseological units specific to social workers’ jargon.

- iv. $X_{l'enfant}$ a des difficultés dans + SN

$X_{the\ child}$ has difficulties in or at + N

- v. difficultés scolaires, maternelles

difficulties at school, maternal difficulties (difficulties with the mother)

- vi. l’enfant (n)est (pas)en (grande) difficulté pour FAIRE qqch

$X_{the\ child}$ is (not) in difficulties to DO sth

- vii. $X_{l'enfant/les\ parents}$ a/ont des difficultés/difficultés de $X_{l'enfant/les\ parents}$ à/pour + FAIRE qqch

$X_{the\ child/parents}$ has/have difficulties to do/not to do sth

When analyzing co-occurrences, a more local phraseology specific to each report considered separately was discovered. 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 were noted at this level of analysis. For instance, in the Thierry-Roux report, a very specific phrase “the father is overcome by his own difficulties,” was used 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 in the other reports as shown in the example x below.

Several unusual phraseological constructions were also uncovered using the co-occurrences methodology. In Thierry Roux’s report example ii, above, the social worker uses the expression “difficulties of the back” instead of the everyday expression “backache”. This unusual usage can certainly be analyzed as an effort (from the social worker) to give to the judge an objective and progressive evaluation. However, this expression can also be interpreted as the result of an automatism. 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 style.

3.3. Repetition and Insistence

Analyzing the frequencies and co-occurrences of the pivot-type *difficulty* led us to go beyond the problematic of phraseology. In fact, a more global strategy built by social workers in their description of what is wrong is progressively put in place. There was a repeated and “invading” presence of the word *difficultés (difficulties)* in reports that are presumed to be very accurate. Is this word systematically linked to words of specification?

In exploring this issue, co-cooccurrence graphs alerted us to an interesting phenomenon in the Thierry-Roux report. The pivot-type *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-type *very* confirm the reciprocity of this “attraction”, figure 6. Moreover, unlike the associative relationships between the pivot-type *difficultés (difficulties)* and its prepositions (figure 5), 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, going back to the text helps understand what is the stake with the use of the word *difficulties* (*difficulties*) in final reports and, more generally, with the use of words that denote “what is wrong” with the child.

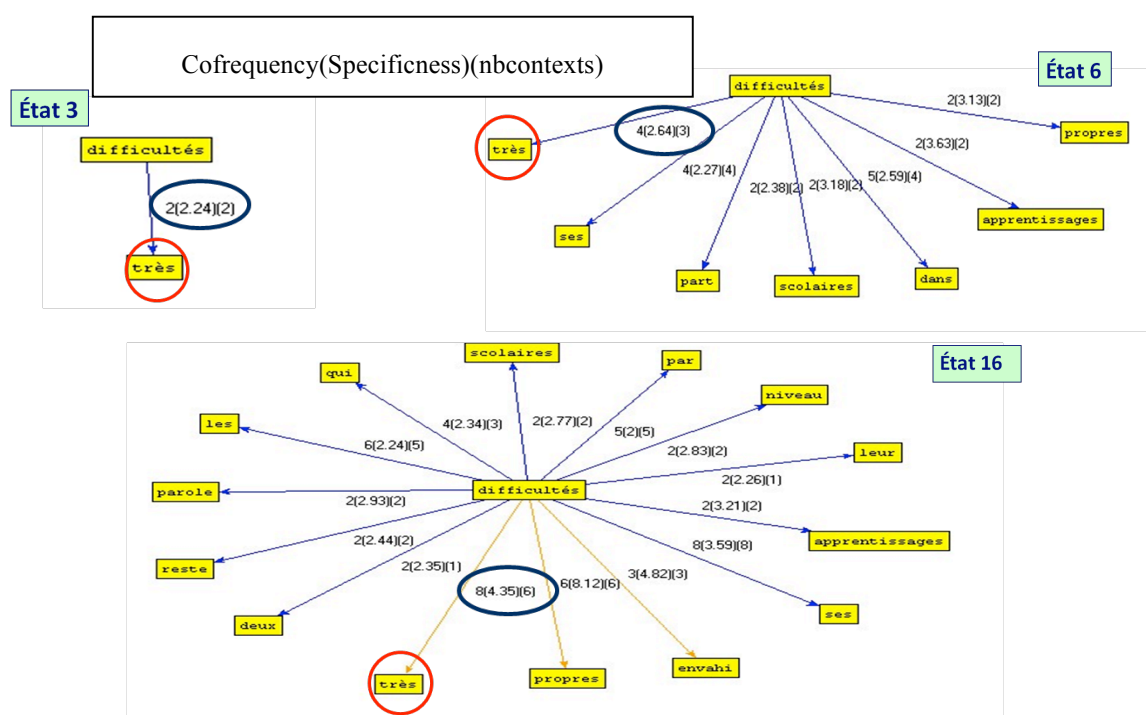


Figure 6- co-occurrences for the pivot-type *difficultés* (*difficulties*) Thierry-Roux drafts 3,6,16, parameters: co-frequency 2, threshold 2

The type *difficultés* 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, these elements of specification are situated in the right co-text of the word *difficultés*, after the naming of what is wrong. Furthermore, when these elements are used in the same sentences as *difficultés*, they often appear in a right-detached position:

- viii. [Thierry-Roux] En troisième, Evelyne a rencontré **des difficultés croissantes**, dans les apprentissages, notamment en mathématiques et en anglais.

*In 9th grade, Evelyne [Thierry's sister] has met **increasing difficulties**, with learning mathematics and English.*

When the writer names parents difficulties, he/she nearly omits systematically giving an accurate description of the problems encountered (two specifications for nine occurrences, and one of the specification in parenthesis). Meanwhile, at this writing stage, the writer always uses markers of intensity: adverbs *very*, *extremely*, adjectives *real*, *important*:

- ix. [Thierry-Roux] Madame est en grande difficulté pour se mobiliser autour des besoins de ses enfants et de l'exercice de son autorité parentale.

*Madam has **many difficulties** to mobilize herself for her children's needs and to exercise parental rights.*

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

- x. [Thierry-Roux] 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 fact, *Mister is very overcome by his own difficulties and while he's claiming visiting rights, he "invests" it as a talking place for himself above all.*

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 on maintaining the placement of the children in example xi.

- xi. 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.

*The relationship of the parents with their children creates **alot of** anxiety for them, each at their level and **in accordance with their own difficulties**. Over the course of the past two years there appears to be an evolution of the four children who remain **extremely** fragile, marked by family behavior that regularly questions their psychological development and necessitates both a **very** important psychological and educative follow-up. It is **fundamental** to maintain the current placement that guarantees the protection of the children.*

This phenomenon is strikingly similar 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 us to reconsider the more global phenomena of repetition and cycle-like behaviour described above. A professional phraseology cannot account for the discourse cycle-like behaviour, repetition and densification of the word *difficulté* observed in these four reports. These phenomena can also be observed in the progressive construction by the social workers of an alert system. In a sensitive situation of communication, this process combines both partial description of "what is wrong" and insistence the problems.

4. Conclusion

The Textometric method sheds light on the inherent cycle-like behavior of discourse allowing us to dynamically reveal strategies used to get social workers' message across or convince a judge of the best possible future for the child being discussed. Co-occurrences made visible these discursive patterns that could have otherwise gone undetected if we were simply to read the reports. This was especially the case with the "insistence" phenomena seen in this paper. Textometry was therefore an appropriate strategy for observing repetition on the entry point, pivot-types *difficulté/difficultés* chosen for this analysis.

The social workers left discursive "clues" to the reader (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 "difficulty" or "problem", "darwer-words" a sort of hyperonym in which they can refer to a variety of non-descript delicate situations. However, these reports are also part of the "job description", meaning they are a routine part of any social worker's 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 reader in this case judge, is expecting, namely an accurate description and evaluation of the situation. Finally, we also think the use of the 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.

Further research using other Textometric methods such as repeated segments (Salem, 1987) and *motifs* (Longrée et al., 2008) will expose other phraseological patterns and their circulation in the discourse structure as well as the logical organization of the corpus. The corpora can also be segmented according to the different sections discussed in part 3.1. This will shift the focus from the progressive variations of the writing process to variations in writing style depending on the topic. The focus of this paper is mostly on observations of repetition, the Textometric method could be applied to studies of substitution and deletion in the writing process. Finally, the corpora could be lemmatized to make the analysis of plural and singular forms (*difficulté/difficultés*) more systematic.

References

- Boutet J., Fraenkel B., Delacambre P. (coord.), (1992). *Les écrits au travail*, Cahier n°6, Journée d'étude du 13 novembre 1992, Réseau Langage et Travail, en ligne sur: <http://www.langage.travail.crg.polytechnique.fr/publications.htm>
- Branca-Rosoff, S (1997). *Modèles de locutionnarité et effets de figement dans le discours politique de l'an II*, in Fiala, P., Lafon, P., Pignet, M.-F. (éds.), *La locution : entre lexicque, syntaxe et pragmatique*, Paris, Klincksieck, p. 285-293.
- Fiala P. (1987). *Pour une approche discursive de la phraséologie. Remarques en vrac sur la locutionnarité et quelques points de vue qui s'y rapportent, sans doute*, Langage et Société, 42, p. 28-48.
- Fiala P., Habert, B., Lafon P., Pineira C. (1987). *Des mots aux syntagmes, figements et variations dans la Résolution générale du congrès de la CGT de 1978*, Mots, 14, p. 45-87.
- Fleury, S., Zimina, M. (2005) (EA2290 SYLED/CLA2T), "mkAlign, Manuel d'utilisation", <http://www.tal.univ-paris3.fr/mkAlign/mkAlignDOC.pdf>
- Fleury, S. (2007). *Le Métier Textométrique: Le Trameur, Manuel d'utilisation*. University Paris 3 Centre de Textométrie. <http://tal.univ-paris3.fr/trameur/>
- Huyette M. (2003). *Guide de la protection judiciaire de l'enfant : cadre juridique, pratiques éducatives, enjeux pour les familles*, Paris: Dunod.
- Lafon, P. (1980). Sur la variabilité de la fréquence des formes dans un corpus, *Mots*, 1 octobre p. 127-165.
- Lafon P., Lefèvre J. (1997). *Le figement : prise en compte discursive, incidences sur les statistiques textuelles et sur l'interprétation*, in Fiala, P. Lafon, P., Pignet, M.-F. (éds.), *La locution: entre lexicque, syntaxe et pragmatique*, Paris, Klincksieck, p. 295-306.
- Lebart, L. and Salem, A. (1994). *Statistique textuelle*. Paris, Dunod, versions auteur disponibles en ligne : à l'ENST <http://ses.telecom-paristech.fr/lebart/ST.html> et à Paris 3 <http://www.cavi.univ-paris3.fr/lexicometrica/livre/st94/st94-tdm.html>
- Lefebvre, J. (2011). *Analyser des écrits promotionnels : prendre en compte les propriétés matérielles de l'espace graphique dans la description grammaticale ?* Linx, 84-94.
- Longrée, D. Luong X, Mellet, S (2008) *Les motifs : un outil pour la caractérisation topologique des textes*, JADT 2008 p 733-744.
- Martinez, W. (2003). *Contribution à une méthodologie de l'analyse des cooccurrences lexicales multiples dans les corpus textuels*, Thèse pour le doctorat en Sciences du Langage, Université de la Sorbonne nouvelle - Paris 3.
- Salem A. (1987). *Pratique des segments répétés*, Paris, Klincksieck.
- Veniard M. (2008). *Ecrire « ce qui ne va pas » dans le champ de l'enfance en danger : les mots problème(s) et difficulté(s)*, in Les carnets du Cediscor 10, Presses Sorbonne Nouvelle.