



Tópicos del Seminario

ISSN: 1665-1200

ses@siu.buap.mx

Benemérita Universidad Autónoma de Puebla
México

Pincemin, Bénédicte
Semántica interpretativa y textometría
Tópicos del Seminario, núm. 23, enero-junio, 2010, pp. 15-55
Benemérita Universidad Autónoma de Puebla
Puebla, México

Disponible en: <http://www.redalyc.org/articulo.oa?id=59415631002>

- ▶ Cómo citar el artículo
- ▶ Número completo
- ▶ Más información del artículo
- ▶ Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

Semántica e interpretación.

Tópicos del Seminario, 23.

Enero-junio 2010, pp. 15-55.

Semántica interpretativa y textometría*

Bénédicte Pincemin

CNRS

Universidad de Lyon

Traducción de Sebastián Giorgi

Introducción

Este artículo es una reflexión y a la vez una exposición de discusiones y de experiencias sobre la pertinencia del enfoque informático textométrico en relación con los principios de la semántica interpretativa.

La textometría se denomina también logometría o estadística textual y es la forma actual de la lexicometría (Lebart *et al.*, 2000). Propone procedimientos de ordenamiento y de cálculos estadísticos para el estudio de un *corpus* de textos digitalizados. La textometría articula sólidamente a estos procedimientos cuantitativos ciertos medios de recorridos y de interpretación cualitativos, determinantes en cuanto a las posibles afinidades con una teoría lingüística como la semántica interpretativa.

* Este texto participa en la reflexión realizada dentro del proyecto *Textometría* (<http://textometrie.ens-lsh.fr/>), por lo cual agradezco mucho a Evelyne Bourion, Carine Duteil-Mougel, Serge Heiden, Sylvain Loiseau, Damon Mayaffre, Céline Poudat y Mathieu Valette por sus atentas y constructivas relecturas, y por los matices importantes y las precisiones que ellos me han permitido aportar al artículo.

La textometría no se confunde con la lingüística de *corpus*, si bien ambas fundan sus investigaciones en un *corpus* digital, cuya constitución es determinante. Como su nombre lo indica, la lingüística de *corpus* persigue un objetivo de descripción y de modelización de la lengua. La textometría, centrada en el texto, ha estado en contacto con diversas ciencias humanas (historia, literatura, ciencias políticas...). Desarrollada en el seno de una comunidad científica interesada en el análisis de datos textuales (ADT),¹ se caracteriza especialmente por algunos cálculos fundadores, estadísticos (las especificidades, las coocurrencias) o no (los segmentos repetidos, las concordancias), y acuerda un lugar primordial al “retorno al texto”, bien previsto por el programa, para interpretar las unidades (generalmente las palabras) seleccionadas por los cálculos. Un estudio que utilice un acercamiento y herramientas textométricos, y que aspire a observar y describir los fenómenos lingüísticos en *corpus*, puede así inscribirse a la vez en la lingüística de *corpus* y en la textometría, sin que ninguna de las dos corrientes subsuma a la otra.

La semántica interpretativa, teoría lingüística desarrollada por François Rastier (Rastier, 1987; Rastier, 2001) nos llevará a recordar los principios cardinales para confrontarlos con las características de la textometría.

Rastier ha sugerido algunas afinidades de su teoría con tecnologías que operan actualmente (Rastier, 1991; Rastier, 2001; Rastier *et al.*, 1994). También explora y experimenta, ya sea en trabajos hechos en colaboración con otros colegas o en la dirección de trabajos de investigación, con algunas herramientas de programas computacionales aplicadas al *corpus*, pero no así para

¹ Las Actas de las Jornadas Internacionales de Análisis Estadístico de Datos Textuales (JADT) publican numerosas comunicaciones sobre cuestiones de teoría textométrica y sobre ejemplos de aplicación variada, así como también acerca de comunicaciones que provienen de otros tipos de acercamiento al análisis textual. Una edición en línea de esas Actas es accesible en el sitio *Lexicometrica*, [disponible en la página electrónica <http://www.cavi.univ-paris3.fr/lexicometrica/>], más especialmente dedicada a la textometría.

la continuación, la profundización y la sistematización de la reflexión sobre las formas de tecnologías apropiadas a un análisis interpretativo. La tesina de Prié (1995) es una de las primeras reflexiones generales sobre este tema. En lo que se refiere al abordaje textométrico,² nos preceden publicaciones que elaboran los usos textométricos en un contexto de semántica interpretativa, generalmente sostenidas por reflexiones teóricas afines a nuestro tema (y estas experiencias nos serán muy útiles aquí), aunque pocas investigaciones se centran en las relaciones entre los fundamentos de la teoría de Rastier y los principios del abordaje textométrico, sino que proponen una reflexión de conjunto, una síntesis: tal es la cuestión que queremos abordar en este artículo.

Nuestro procedimiento será en cuatro tiempos. Desde un principio hay que dejar atrás los *a priori* negativos sobre la textometría, en cuanto a su compatibilidad con un acercamiento lingüístico. Luego, el artículo invita a compartir el entusiasmo de los primeros descubrimientos —primeros encuentros luminosos de la semántica interpretativa con la textometría— y su consolidación en una proposición metodológica para el análisis temático. El tercer tiempo es la parte medular de este artículo: tratar de comprender plenamente las afinidades de la teoría semántica y de las herramientas provistas para el análisis, volviendo a sus principios fundamentales, y tomar distancia con una visión englobante. Sobre esta base, el cuarto y último tiempo puede abrir algunas perspectivas prometedoras.

1. Discusión de reducciones a primera vista comprometedoras

1.1. *El texto, ¿un “saco de palabras”?*

Para la textometría, el *corpus* está segmentado en unidades (habitualmente del tamaño de las palabras): eso está relacionado

² Otras tecnologías pueden aportar de manera complementaria a los análisis de *corpus* en semántica interpretativa (*cfr.* nota 31), y otras teorías lingüísticas pueden inspirar y guiar benéficamente algunas prácticas textométricas.

con el procedimiento técnico de indexación, necesario para la aplicación eficaz tanto de las funcionalidades de búsqueda de motivos como de cálculo estadístico. El *corpus* puede estar estructurado en partes, típicamente considerados textos. En ciertos programas computacionales³ es igualmente posible definir las vecindades locales utilizables para las búsquedas o cálculos de coocurrencias. Primera observación positiva: el texto no está tan desestructurado como se podría entender con la imagen de “saco de palabras”: La textometría moviliza (i) una contextualización global, (ii) una contextualización local, y registra (iii) la relación de orden de las palabras según la linealidad del texto. Se trata, de hecho, de no atribuir al conjunto del *análisis* textométrico lo que es momentáneamente requerido para *un cálculo*: así, un cálculo estadístico de especificidades moviliza fuertemente las contextualizaciones globales, pero ignora las relaciones de sucesión inmediata de palabras;⁴ otro cálculo, de coocurrencias por ejemplo, puede considerar sólo las vecindades locales e ignorar tanto el encadenamiento exacto de palabras como su contextualización global; un tercero, de segmentos repetidos, sólo considerará los encadenamientos de palabras, sin consideración para su contextualización global. No obstante, los resultados de estos cálculos no deben interpretarse sin ayuda de metódicos retornos al texto, que permiten leerlo y observar directamente las vecindades, sucesiones y localizaciones completas de palabras.

Las investigaciones actuales en textometría son particularmente sensibles al desarrollo de nuevos tratamientos que integran la dimensión sintagmática del texto: “topología”, basada

³ Las frases en Weblex (explotadas por el motor de búsqueda CQP), las secciones de Lexico 3, etc.

⁴ El recorte en partes puede también ser un medio de restituir, en una primera aproximación, el desarrollo sintagmático del texto (la “táctica”, en términos de la semántica interpretativa), o el desarrollo cronológico del *corpus*. Por ejemplo, Bourion (2001) propone un estudio de *Père Goriot* (*Papá Goriot*) estructurado por su recorte en capítulos.

en una modelización del desarrollo en series sucesivas; “topografía”, con el éxito⁵ del mapa de secciones implementado en Lexico 3 (Mellet y Salem, 2009).

Además, no es porque en un momento dado el cálculo se apoye en *una* segmentación en unidades y *un* recorte en partes englobando esas unidades, que la elección de esas unidades y esas partes en textometría sea única y predefinida.⁶ Si la opción más corriente consiste en estudiar el *corpus* a través de su léxico, también se lo puede ver por la vía de otras descripciones (p. ej. categorías gramaticales) y de otros planos (el morfema o su aproximación por los “tri-grams”, alternancia de palabra, etc.). Asimismo, la variación en la elección de la división (recorte del *corpus* en partes) es una forma de dar cuenta de categorías metatextuales y de caracterizaciones filológicas (como el género, el autor, el período histórico), de variables situacionales significativas de la producción de textos, como de estructuraciones intratextuales (p. ej. capítulos).

1.2. Simples cadenas de caracteres

El “saco de palabras” tendría no sólo la debilidad de ser un saco, sino también la de no reunir incluso generalmente palabras (definidas de manera lingüísticamente aceptables), simplemente unas cadenas de caracteres extraídas de modo mecánico del *corpus*. De manera más general, se trata primero de una constante pragmática: en la práctica, el análisis se basa en el material tex-

⁵ Aquí no hay un cuadro sinóptico científico, pero podemos constatar fácilmente que los mapas de secciones son, por ejemplo, copiosamente movilizados en la selección de estudios (Salem y Fleury, 2008), reuniendo ejemplos de análisis diversificados realizados con el programa Lexico 3.

⁶ Incluso si en ciertos programas la importación del *corpus* supone la elección de una sola segmentación en “palabras” (p. ej. Lexico 3) o de un sólo recorte en “partes” (p. ej. Hyperbase), nada impide crear tantas bases textométricas —tantas versiones interrogables del *corpus*— como uno desee, variando la definición de unidades o de partes.

tual digitalizado, pues aun si la apuesta es acceder a observaciones semánticas “se manipula las cadenas de caracteres para estudiar los significados” (Rastier, 2001: III.1.2).

La textometría no pretende partir de significantes lexicales; antes bien, se basa en una segmentación lingüísticamente simple, aproximativa.⁷ Dicho de otra manera, relativiza la cuestión de la buena elección de las unidades iniciales.⁸ Al hacer esto, adopta, según nosotros, una posición muy afín con la semántica interpretativa: porque si tolera la posible simplicidad de la representación inicial⁹ es porque ella considera que las verdaderas unidades no se hallan al inicio sino al final de los tratamientos. Por la visión global que ellas integran, las estadísticas y otros recuentos dan muestra de las grandes regularidades que atraviesan el *corpus* y determinan así la redefinición de las unidades locales. Concretamente, por ejemplo, es así como la técnica de segmentos repetidos ha sido pensada desde los comienzos de la textometría (Lafon y Salem, 1983): rectificar y ajustar *a posteriori*, dado el *corpus*, segmentaciones iniciales inconvenientes.

Algunas experiencias vertiginosas han demostrado también, en los hechos, la sorprendente robustez de los análisis estadísticos.

⁷ Las segmentaciones clásicas propuestas en los programas computacionales de textometría están definidas directamente en la cadena de caracteres, sobre criterios tipográficos, desde luego, precisos y pertinentes pero no siempre en concordancia con las estructuras lingüísticas. Evidentemente, si el *corpus* está Enriquecido y presenta un análisis en unidades, éste en principio puede explotarse por el análisis textométrico, en remplazo de, o en complemento de, una segmentación de orden tipográfico (esto depende de la implementación del programa computacional).

⁸ A pesar de todo, es verdad que la comunidad textométrica ha estado durante mucho tiempo atravesada por los famosos debates sobre la *lematización* previa de los *corpus*, ahora poco a poco superada con la coexistencia de la articulación de varias descripciones (Mayaffre, 2005).

⁹ Sin embargo, no necesariamente es apropiado considerar todo análisis inicial (Poudat, 2006). En particular, un análisis irregular, de difícil interpretación (opacidad de su mecanismo de producción o de la significación efectiva de las etiquetas) o no adaptado al *corpus* y a la objetividad de la investigación compromete el rendimiento de los cálculos textométricos.

Brunet (2006b) imagina describir el *corpus* no solamente por sus grafías (las palabras en tanto que cadenas de caracteres entre dos blancos u otros caracteres separadores) o sus lemas, sino también a partir sólo de sus etiquetas morfosintácticas (p. ej. *ncms* para *nombre común masculino singular*, independientemente de la forma de este nombre), por las secuencias de categorías gramaticales entre dos puntuaciones (p. ej. *pvdn* para *pronombre verbo determinante nombre*), por las palabras descompuestas en secuencias de cuatro caracteres (p. ej. *ventana*, siendo representada por *vent*, *enta*, *ntan*, *tana*),¹⁰ por las grafías reescritas como sucesiones de consonantes y vocales (p. ej. *tanto* como *santo*, *harto*, *largo*, etc. devienen CVCCV).¹¹ Ahora bien, globalmente, un análisis factorial o arborescente aplicados a estos textos, representados de maneras tan diversas y a veces reductoras,¹² hace destacar las mismas configuraciones de proximidades o de oposiciones entre sí. Dicho de otra manera, independientemente de qué tan gastada y pobre sea la descripción inicial, ciertas regularidades textuales son tales que pueden ser captadas por un tratamiento que explote plenamente la dimensión global del *corpus*.

1.3. Eliminación frecuente de los hápax, que podrían ser lugares de singularidades significativas

Para aligerar los tratamientos estadísticos, un umbral mediante las frecuencias es practicada comúnmente: por lo general, las palabras de débil frecuencia —y también las palabras gramaticales de alta frecuencia— son apartadas del cálculo. En el caso

¹⁰ Ejemplo del texto original “ex. *fenêtre* étant représenté par *fenê*, *enêt*, *nêtr*, *être*”. [N. del T.].

¹¹ Ejemplo del texto original “ex. *tant comme sont, dans, sang...* deviennent CVCC”. [N. del T.].

¹² Hasta el sentimiento de una representación desnaturalizada: “desencarnación” del texto, “datos pervertidos”, “pérdida [...] irremediable” (Brunet, 2006b).

particular de los hápax (palabras de frecuencia 1) es incluso naturalmente favorable, en ocasiones, separarlos, ya que son de entrada no pertinentes si el cálculo se apoya sobre las repeticiones. El umbral sobre las frecuencias permite el análisis, con una rauda mirada, a las estructuras de conjunto. Sin embargo, por experiencia, no debería ser la única mirada sobre el *corpus*, y exploraciones o cálculos más focalizados son la ocasión de reconsiderar las unidades apartadas en un primer tiempo.

No obstante, no es seguro que las singularidades pertinentes para la descripción lingüística correspondan necesariamente a las bajas frecuencias en el plano estadístico. En el proyecto Princip (Valette, 2004) se fija la atención en las neologías, que en un principio suelen ser escasas, pero están, de hecho, compuestas de morfemas que pueden estar, por el contrario, muy presentes en el *corpus* y que se manifiestan como mejores unidades de caracterización que el léxico.¹³ O incluso, detrás de la variedad de manifestaciones de un tema (Rastier, 1995) —lexicalizaciones diversas, sintéticas o difusas, etc.— podríamos encontrar una modelización única y, por lo tanto, con bastante frecuencia solicitada bajo la forma de un conjunto de palabras “isotopantes” (que manifiestan juntas un sema común), y reallizándose por la coocurrencia de algunas de ellas (*cfr.* el concepto de *comunidad* en Bommier-Pincemin, 1999). Recurrir a un diccionario sémico sería una técnica inversa para captar y amplificar las recurrencias de semas (Reutenauer *et al.*, 2009). En pocas palabras, los elementos sobre los cuales están fundadas las descripciones no son quizá tan raros ni únicos como pudiera parecer, incluso si las manifestaciones aparentes en el *corpus* son singulares.

¹³ El trabajo en el nivel morfológico es también muy pertinente para el análisis de la terminología científica, con los procedimientos de conceptualización y de préstamo (Loiseau, 2006; Valette, 2006).

1.4. Tratamiento cuantitativo vs. cualitativo

Los cálculos textométricos son, por supuesto, cuantitativos. Ahora bien, la semántica interpretativa no es una semántica formal en la cual el sentido se modelizaría como un cálculo. No obstante, un acercamiento cuantitativo puede encontrar una pertinencia si el volumen del *corpus* es consecuente.¹⁴

En primer lugar, la textometría explota especialmente los modelos estadísticos. Se mide la distancia entre una repartición aleatoria de palabras y su comportamiento efectivo. El principio operatorio consiste entonces en considerar la lengua, y más precisamente su uso manifestado en *corpus*, como regulada por limitaciones lingüísticas, opuesto al azar (Lafon, 1980). Observamos, en efecto, que los cálculos hacen reaparecer vínculos lexicales, sintácticos, semánticos (isotopías de la semántica interpretativa),¹⁵ genéricos (propio del género textual), estilísticos...

Si los cálculos son cuantitativos, el análisis textométrico integra una gestión cualitativa englobante. Antes de proceder al cálculo, se trata en primer lugar de formular una problemática de búsqueda, de construir un *corpus* pertinente, de determinar un punto de entrada apropiado y de variar eventualmente los sub-*corpus* de trabajo, de definir el tipo de tratamiento adaptado,

¹⁴ El lector humano es evidentemente el mejor intérprete de un texto, en comparación con todo tratamiento automático y mecánico, que nunca es una verdadera lectura. Las ventajas de la computadora son su rapidez de cálculo y su memoria: el interés está en aprovecharlos para ayudar la lectura humana, sugerirle puntos de apoyo y pistas de investigación cuando el volumen de los textos supera las capacidades cognitivas.

¹⁵ “El concepto de isotopía está [...] basado en la noción de redundancia de la información, es decir, en cierta manera, sobre un elemento cuantitativo. Si los rasgos de una isotopía no son directamente observables, puesto que son elementos del significado y no del significante, el carácter cuantitativo puede ser la base de su identificación. Los conceptos descriptivos de la semántica interpretativa no son, por lo tanto, desarrollados para un marco metodológico cuantitativo, pero ofrecen muchos puntos de articulación para la interpretación de datos cuantitativos” (Loiseau, 2006: 30).

de ajustarlo, llegado el caso: en resumen, se trata de formular diversas operaciones cualitativas determinantes. Y después del cálculo, es bien sabido que lo que se obtiene es un resultado, no una respuesta.¹⁶ Falta toda la parte de interpretación, de cualificación eventual de ciertos fenómenos, y de la progresión del análisis al elaborar un recorrido interpretativo.¹⁷ Nada de eso es un dato en sí, ni tampoco el resultado de un cálculo. La automatización del cálculo no condiciona en absoluto a un uso mecánico.

2. Experiencias positivas reveladoras

Luego de haber apartado las objeciones que impiden considerar incluso el abordaje textométrico, invitamos al lector a compartir los primeros descubrimientos de la textometría efectuados por la semántica interpretativa.

2.1. El contraste de un texto en relación con un corpus de referencia

Es necesario comenzar por mencionar la experiencia memorable, incluso para el mismo Rastier, de ver cómo el cálculo de la diferencia-reducida pone en evidencia, en el texto de un relato de Maupassant, formas claves para su interpretación:

[...] necesité diez años para comprender la importancia del número *diez* en la novela de Maupassant titulada *Toine* (cf. el autor, 1989, lib. II, cap.V); en cambio, el test de la diferencia reducida [aplicada en el

¹⁶ El cálculo siempre proporciona un *resultado*, aunque éste sea vacío; sin embargo, si está mal concebido o no entra en una gestión de análisis metodológico que le da sentido, el resultado no aporta ninguna *respuesta*, tan sólo unos resultados ociosos y en general voluminosos.

¹⁷ “[...] lo cuantitativo y lo cualitativo no se oponen de ninguna manera: sólo un análisis cualitativo puede hacer significativos los fenómenos cuantitativos notables” (Rastier, 2001: VII.3.5).

marco de un *corpus* de referencia cuidadosamente constituido] me lo puso instantáneamente ante los ojos y aun me permitió sacar provecho de una ocurrencia en la primera línea que, lo confieso, se me había escapado, a pesar de que ella hubiera reforzado mi propósito (Rastier, 2001: III.3.5 nota 1).

El cálculo lo había producido Bourion, con la ayuda de programas que ella había concebido junto con Maucourt.¹⁸

ON le connaissait à DIX lieues aux environs le père Toine, le GROS Toine, Toine-ma-FINE, Antoine Mâcheblé, dit Brûlot, le CABARETIER de Tournevent.

Il avait rendu célèbre le HAMEAU enfoncé dans un pli du vallon qui descendait vers la mer, pauvre HAMEAU PAYSAN composé de DIX MAISONS normandes entourées de fossés et d'arbres.

Elles étaient là, ces MAISONS, blotties dans ce ravin couvert d'herbe et d'ajonc, [...]

Figura 1

Principio del relato *Toine* de Maupassant, destacando en mayúscula unas grafías específicas (diferencia-reducida >3, frecuencia en *Toine* >2) en relación con un *corpus* de referencia de 350 novelas y relatos de 1830 a 1970 resultado de la base de datos Frantext (*corpus* descrito en Bourion, 2001).

¹⁸ Aunque al abrir la parte del artículo consagrado a la historia del descubrimiento de la textometría en el marco de la semántica interpretativa, la experiencia de Rastier no es en rigor inaugural, ya que Bourion había emprendido ya de manera significativa una reflexión y algunas experimentaciones sobre el tema. Pero esta experiencia puede considerarse como una etapa que estimula y refuerza las investigaciones en el dominio.

2.2. La coocurrencia al servicio de la descripción temática y semántica

Luego vino rápidamente una puesta en relación de conceptos teóricos de isotopía (recurrencia de un sema) y de molécula sémica (agrupamiento estable de semas) con un cálculo de coocurrencia, experimentado inicialmente en el marco de estudios temáticos sobre los sentimientos en la novela francesa (Rastier, 1995); luego retomado en otros contextos (Deza, 1999; Bourion, 2001; Valette, 2004; Poudat, 2006; Loiseau, 2006). Los pasos metodológicos se sintetizan así:

Resumamos las principales etapas de una investigación temática asistida: (i) Elección de las hipótesis en función del objetivo general de la investigación (un pre-análisis estadístico puede guiar la investigación de las hipótesis, pero la frequentación previa del *corpus* es indispensable para guiar las intuiciones). (ii) Investigación de los co-ocurrentes por el método estadístico de los diferenciales reducido o hipergeométrico. (iii) Transformación interpretativa de los co-ocurrentes en correlatos y constitución de las redes temáticas (esta etapa es facilitada si se practica una interrogación simultánea en varios co-ocurrentes;¹⁹ cf. Bourion, 1995: I.2). (iv) Convalidación de los resultados por cruce del análisis temático con el análisis de otros componentes del mismo *corpus*, pro prueba en un *corpus* de control o por confrontación con otras investigaciones temáticas (Rastier, 2001: VII.3.4).

Lo que produce el cálculo textométrico es, entonces, los coocurrentes, en el plano de los significantes; y lo que se pretende es la obtención de correlatos, en el plano de los significados. Pasamos de los primeros a los segundos por una interpretación que reconoce la presencia de un rasgo semántico común entre la o las palabras que sirven de inicio de la investigación, y el coocurrente entonces calificable como correlato.

¹⁹ Los coocurrentes son elegidos en función de hipótesis sobre la variación de lexicalización de los rasgos por observar.

Entre los coocurrentes, particularmente los coocurrentes a distancia reducida, se mezclan a los correlatos de palabras que están en relación fraseológica: esto se ha observado en particular para el léxico de partes del cuerpo (para *cœur* —corazón—: *avoir à cœur* —importarle mucho algo a alguien—, *savoir par cœur* —saber de memoria—, etc.; para *pied* —pie—: *de plain pied* —de una sola planta—, *faire le pied de grue* —estar de plantón—, *sur un pied d'égalité* —en un pie de igualdad—, etc.). Pero eso puede no ser tan negativo: por un lado, las locuciones son tanto más fácilmente reconocibles e interpretables como el fenómeno es ahora bien conocido; más sutilmente, y en particular en ciertos *corpus*, siempre es posible la desarticulación de una locución que vuelva a motivar semánticamente estos componentes sueltos. Ahora bien, ciertos índices podrían ayudar a distinguir las fraseologías y los correlatos posibles —considerable valor de correlación, en particular para las formas no *lematizadas*, y posicionalmente orientado— (cfr. Bourion, 2001: 58); y la búsqueda de coocurrentes debe, pues, privilegiar un contexto no demasiado estrecho, del tamaño del párrafo (Rastier, 2001: VII.3.4; Deza, 1999: cap. 5).

En las investigaciones de semántica interpretativa estos cálculos de coocurrentes han sido equipados principalmente por dos programas: un programa informático desarrollado en el INaLF por Maucourt, y la función Tema en el programa *Hyperbase* de Brunet. El programa *Hyperbase* (Brunet, 2006a) es uno de los más difundidos en textometría, y, luego de su introducción, la función *TEMA* ha confirmado su pertinencia. El programa de Mancourt merece aquí una exposición rápida, ya que no ha conocido la misma difusión, y sobre todo ha sido desarrollado en el contexto de investigaciones explícitamente dirigidas en el marco de la semántica interpretativa, en una colaboración del informático con Bourion. Es movilizado repetidas veces en el volumen (Rastier, 1995), y encontramos una presentación de su forma más lograda en la tesis (Bourion, 2001). Este programa determina la lista de coocurrentes estadísticamente

significativos de una palabra polo, según la medida de la diferencia-reducida.²⁰ Presenta el resultado bajo la forma de lista (organizada y ordenada en función del valor de la diferencia-reducida y de la frecuencia), pero también fija los contextos de coocurrencia bajo la forma de concordancia, poniendo tipográficamente de relieve los coocurrentes, y seleccionando las líneas de contexto en función de los coocurrentes.

| | | |
|--|---|-----------------|
| ils s' ABATTIRENT , haletants , au | pied d'un BUISSON incendié par les rayons du soleil couchant | MAUP.cn 1881 |
| I' OMBRE s' ABATTAIT inerte et lourde au | pied des ORMES | FRAN.om 97 |
| lie dernière de la clientèle conquise , ABATTUE aux | pieds du tentateur , évoquait l' image fière et VENGERESSE de | ZOLA.BD 1883 |
| PRêTRE se baissa vers *suzanne , toujours ABATTUE au | pied du LIT , la releva , la MIT dans un | MAUP.cn 1883 |
| d' ACAJOU , assez profondes , supportées sur deux | pieds de BRONZE et remplies d' une foule de ces beaux | SUE.A-G1831 |
| vers le soir , *kai - *koumou ACCOSTA au | pied des MONTAGNES dont les premiers CONTREFORTS TOMBAIENT à | VER.ecG 1868 |
| n' ACCOSTA le rivage qu' à plusieurs milliers de | pieds de l' ENDROIT qui faisait face au point d' où | VER.im 1874 |
| les vêtements étaient ACCROCHÉS au | pied du LIT , sous la moustiquaire . | MALR.ch 1933 |
| . ACCROUPIS au | pied d' un MUR , trois hommes mangeaient du pain qu' | PER.c 1965 |

²⁰ La diferencia-reducida produce un valor numérico que mide el carácter no aleatorio de su coocurrencia. Se selecciona entonces una lista de coocurrentes significativos, conviniendo un umbral sobre el valor absoluto de la diferencia-reducida.

| | | |
|---|--|-----------------|
| . le gamin , ACCROUPI au | pied du PARAPET , s'affairait à manier sa pelle , | MART.Te 1940 |
| le soleil ou sous la pluie , ACCROUPIS au | pied d'un SAULE , le cœur battant , l'âme | MAUP.cn 1886 |
| de la GRANDE CASE des douanes , ACCROUPIE au | pied d'un manguier , une négrillonne gémisait . | MILLE.B 1908 |

Figura 2

Ejemplo de documento producido por el programa de Maucourt: contextos de “pie(s) de” que contienen varios coocurrentes seleccionados por la diferencia-reducida, y ordenado alfabéticamente sobre el coocurrente de izquierda más próximo (Bourion, 2001, tomo II: 8).

El trabajo de descomposición de tales extracciones de contextos se organiza reagrupando los contextos que realizan el mismo motivo semántico, el mismo tema, definible abstractamente como ciertos semas estructurados de manera actancial:

Cuando estudiamos los coocurrentes (plano de la expresión) para calificarlos eventualmente de ‘correlatos’ del tema (plano del contenido), localizamos igualmente las relaciones casuales, lo que permite representar el tema como un grafo donde los nudos representan los componentes y las relaciones los primitivos (ergativo, acusativo, atributivo, dativo, benefactivo, instrumental, final, *cfr.* Rastier F., 1989: 62-65) (Bourion, 2001: 116).

La lectura de los contextos descriptivos de personajes y la localización de rasgos característicos comunes permite también abstraer sus roles bajo la forma de agonistas, es decir, un trabajo no solamente sobre el componente temático del texto sino también sobre el componente dialéctico, en el sentido de la semántica interpretativa.

Los coocurrentes, como índices potenciales de isotopías (Mayaffre, 2008) pueden también aprovecharse para contrastar semánticamente los lexemas *a priori* próximos. De esta manera, Deza (1999) estudia los coocurrentes respectivos de *piedad*, *commiseración*, *compasión* y *misericordia* en un *corpus* de novelas para caracterizar mejor su sentido efectivo en ese contexto. Loiseau (2006) muestra igualmente cómo el estudio de un concepto a través de sus coocurrentes (p. ej. *naturaleza* en la obra del filósofo Deleuze) puede ayudar a delimitar a la vez su unidad de sentido y sus diversas acepciones.

2.3. Hacia la localización de pasajes

Bourion (2001) sugiere una prolongación en la localización de correlatos temáticos a través de una presentación de resultado más selectiva y más flexible que las líneas de concordancia del programa Maucourt:

Queda por concebir programas que busquen automáticamente partes de textos que comporten un número importante (y estadísticamente pertinente) de palabras relacionadas con el miedo (el campo lexical de estudio), y también de aquellas que hemos calificado como correlatos: Unas ‘ráfagas’ señalarían pasajes probablemente indexados sobre la isotopía del miedo (Bourion, 2001: 106).

Existe un cálculo textométrico llamado “ráfagas”, pero es más bien del lado de la función “Frases-clave” de *Hyperbase* (Brunet, 2006a) adonde habría que encontrar una primera realización de esta idea. Los lingüistas de la semántica interpretativa conocen poco esa función, y por lo demás, sería posible revisar su concepción (el cálculo ha sido realizado por una sucesión de ajustes heurísticos, sin haber tenido aún la ocasión de un debate científico²¹). Habría, sin embargo, en la concepción de tal

²¹ Se asemeja, no obstante, al cálculo de respuestas modales expuesto en (Lebart *et al.*, 2000).

funcionalidad un espacio de colaboración entre la semántica interpretativa y la textometría.

Otra forma de localización de pasajes propuesta por la textometría consiste no en una extracción selectiva, sino en una representación gráfica del conjunto del texto, con un indicador de densidad de correlatos en el transcurrir del texto. Ehrich (1995) dibuja así unos grafos que figuran las manifestaciones del tema *ambición* en *Papá Goriot* (Balzac). La textometría ha concebido luego otra representación adaptada a la percepción de fuertes densidades de ocurrencia o de coocurrencias en el seno de un *corpus*: el mapa de secciones (Lamalle y Salem, 2002). Pero la cuestión del indicador que estaría por construirse para medir esta densidad (integrando consideraciones no solamente de frecuencia sino también de diversidad, de especificidad, etc.), y la cuestión más delicada de la delimitación de un pasaje queda por trabajar, tanto en el plano teórico como en el plano técnico.

De hecho, los desarrollos ulteriores de la semántica interpretativa han precisado el concepto de *pasaje* y señalado su importancia fundamental en la descripción lingüística, reconociéndola como *signo* que articula los planos del significado y del significante. Ahora bien, la localización de zonas densas en coocurrentes podría ser un punto de apoyo para la localización de pasajes:

Cuando se apoya sobre *corpus* de textos que pertenecen al mismo género y al mismo discurso que el texto analizado, el test de la diferencia reducida permite localizar agrupamientos de coocurrentes que son buenos candidatos para la constitución de pasajes (Rastier, 2008).

2.4. La caracterización de textos y de géneros textuales

El análisis temático mediante los coocurrentes puede ser puesto al servicio de la caracterización de textos. En el proyecto Princip (Valette, 2004) lo que está en juego es localizar y discriminar

las páginas racistas y las páginas anti-racistas en internet. La textometría se ha movilizado para construir temas específicos para unos y para otros: para una lexía que pertenezca a un fondo isotópico común, como “inmigración” o “extranjero”, se buscan sus correlatos en el sub-*corpus* racista y en el anti-racista. Los sub-*corpus* se revelan también caracterizables por índices de toda naturaleza, no solamente lexicales sino también morfológicos y semióticos.

La textometría ha permitido también observar concretamente la incidencia semántica del marco genérico, que ocupa un lugar determinante en semántica interpretativa. Así por ejemplo, en un *corpus* de literatura francesa se ha evidenciado que la palabra “amor” no atrae los mismos correlatos y, por lo tanto, no construye las mismas isotopías ni los mismos temas si se encuentra en novelas o en poesías (Bourion, 2001: 42-47). Del lado de la novela se señalan contextos de uso tales como: *amor platónico, inspirar (el amor/un amor...), amor exclusivo*; e incluso, *pasión, apasionado, celos, ambición, dolor, orgullo, olvido, renunciamiento, revulsión, vanidad, egoísmo*. Y para la poesía: *alegría, himen, himeneo, ígneo, sonrojar, trampa, ruißeñor, suspiro*, entre otras.

Los análisis factoriales sobre los recuentos y medidas proporcionados por la sociedad Synapse en su *corpus* (Malrieu y Rastier, 2001; Beauvisage, 2000), han sido igualmente experiencias concluyentes que confirman la determinación de lo global sobre lo local, y las interrelaciones transversas a los planos de descripción. Estas experiencias de lingüística de *corpus* están, sin embargo, al margen de la textometría, pues incluso si el análisis factorial es un cálculo practicado corrientemente en textometría, la ausencia aquí de retorno al texto (la sociedad Synapse no da acceso a sus fuentes) no permitía hablar propiamente de una gestión textométrica. Por el contrario, y con el mismo espíritu, pero esta vez controlando el análisis mediante retornos al texto, Poudat (2006) procede a la descripción de un género textual, el del artículo científico de

lingüística francesa, según los principios de la semántica interpretativa y recurriendo a procedimientos textométricos disponibles en el programa DTM.²² La descripción se apoya tanto en un análisis lexical como en una descripción morfosintáctica, y explora diferentes planos, infratextuales (como la sección) y supratextuales (el estilo del autor, el dominio...).

Para describir los textos y los géneros, la semántica interpretativa propone interesarse en el componente temático, pero también en otros tres componentes (dialéctico, dialógico, táctico). Tomar en cuenta estos componentes diferentes es tan importante como su funcionamiento en interacción. Las tesis de Loiseau (2006) y Poudat (2006) proponen nuevas maneras de movilizar los cálculos textométricos en el espíritu de la semántica interpretativa, especialmente los diagramas de distribución en diferentes planos, como aquel del texto o del párrafo (“diagramas tácticos” y “gamas de densidad”) (*Cfr.* Loiseau, 2006: §12.F.).

3. Connivencias de fondo

Las experiencias precedentes han mostrado el carácter prometedor del abordaje textométrico para una investigación sobre el *corpus*, en el marco de la semántica interpretativa. Para confirmar la intuición, nos proponemos ahora relevar de manera más completa las adecuaciones esenciales entre la teoría de la semántica interpretativa y los principios fundadores del método textométrico.

²² El software DTM está concebido y desarrollado por Lebart, y está difundido en la dirección <http://www.dtm.vic.com/>. A dominancia estadística, está especializado en los procedimientos de análisis de datos (análisis factorial, clasificación...) y en las técnicas matemáticas de ayuda a la interpretación de resultados.



3.1. Lingüística y semántica

La semántica interpretativa se interesa en el sentido. Según ella, el sentido puede elaborarse a partir de índices morfológicos, sintácticos, fonéticos, etc. Puede ser transversal a las categorías gramaticales, movilizar de igual manera una puntuación, un aspecto verbal, un morfema, un perfil prosódico y rítmico, una tipografía o una compaginación (Bourion, 2001; Malrieu y Rastier, 2001; Beaudouin, 2002; Valette, 2004; Loiseau, 2006). La textometría está en condiciones de tomar en cuenta descripciones de textos de toda naturaleza, incluso si están poco explicitados por una codificación del *corpus*.

El enfoque semántico propuesto por la textometría está en total acuerdo con la demanda de “de(s)-ontología” formulada por Rastier.²³ Se trata de evitar todo presupuesto reductor; queremos, sobre todo, permanecer lo más cerca del texto y no comenzar por estudiarlo a través del prisma de una ontología. Así, desde sus comienzos, la textometría se convierte en una especialidad del tratamiento de las preguntas abiertas en las encuestas para evitar la post-codificación de las respuestas (entre la encuesta y el análisis) que borra variaciones de expresiones potencialmente significativas (Lebart *et al.*, 2000).²⁴ De igual manera, los programas computacionales de análisis textual que proponen tratamientos estadísticos o cuantitativos algunas veces

²³ En el plano experimental, Deza (1999) muestra especialmente cómo la canonicidad que se expresa en el *corpus* está desfasada con una representación puramente referencial del mundo, con el ejemplo de la edad de los personajes en la novela francesa.

²⁴ Un ejemplo dado por (Lebart y Salem, 1994: 169 y 188) a la pregunta: “¿Cuáles son las razones que, según usted, pueden hacer dudar a una mujer o a una pareja de tener un hijo?”, las respuestas “falta de dinero” y “razones financieras” (o incluso “el trabajo de la mujer” y “la mujer trabaja”) serían *a priori* asimiladas a la misma respuesta por una post-codificación; ahora bien, un estudio textométrico que preserva estas formulaciones muestra que estas maneras de expresarse son significativamente correlativas a encuestados de categorías socio-profesionales muy contrastadas, y donde se reconocen matices importantes.

muy próximos a los tratamientos textométricos, pero que comienzan por remplazar el texto por una representación en términos de categorías predefinidas (proyectando el texto sobre una ontología), salen claramente del método textométrico.²⁵

La preocupación sobre la fidelidad al texto se ha expresado vivamente incluso en el debate que atraviesa la comunidad textométrica y que concierne a la *lematización*: ¿Hay que segmentar, en verdad, el texto en formas gráficas “tal cuales”? , o bien ¿no es más justo aplicar un pre-análisis puramente morfosintáctico que asimile todas las formas flexivas de una misma palabra en la entrada del diccionario correspondiente? Dicho de otra manera, ¿elegimos contar y analizar independientemente *flor* y *flores*, *es*, *éramos* y *será*? , o bien ¿preferimos reconocer aquí sólo los lemas *flor* y *ser*? La cuestión ha quedado largo tiempo abierta, pues si la *lematización* era seductora para desambiguar eficazmente numerosas homografías (p. ej. Un *partido* político vs. *he partido*), la atención al texto había revelado también que a menudo las diferentes flexiones eran portadoras de una semántica diferente, los plurales eran típicamente más concretos que los singulares (p. ej. *le travail* —*el trabajo*— vs. *les travaux* —*las obras*, *por ejemplo*, *públicas*—) (Geoffroy, Lafon, Tournier, 1974). La semántica interpretativa comparte la sensibilidad de no aplastar estas distinciones, como lo muestra el estudio de Bourion sobre *al pie de* vs. *a los pies de* en un *corpus* de literatura francesa: el singular reenvía a descripciones de localización, con un sema de /verticalidad/ (*al pie de la montaña*, etc.), mientras que la forma plural corresponde a escenas de

²⁵ Entre los softwares que proponen un análisis textual vía una reducción a categorías, y no trabajando sobre el texto mismo, podemos citar Tropes (que sin embargo implementa unos cálculos de origen textométrico como los ráfagas) o el análisis semántico liberado por Cordial (mientras que el componente analizador morfo-sintáctico del mismo software puede muy bien preparar un *corpus* para un análisis mediante un software textométrico). El caso de Prospero es intermedio, en el sentido de que las categorías están construidas por el usuario —uno no está sometido a un diccionario “universal” predefinido.

imploración (*arrojarse a los pies de alguien*), apelando esta vez a los semas /humano/ y /sentimiento/ (Bourion, 2001: 62).

3.2. Semántica de los textos

3.2.1. El lugar central de los textos en todas las etapas del análisis textométrico

El primer objeto empírico de la lingüística no es la frase o la proposición sino el texto. La realidad observada es en principio la de los textos, situada en las prácticas, y no en las frases recibidas independientemente del texto del cual son resultado. En los términos de la semántica interpretativa, desarrollada luego en semántica de los textos, lo global determina lo local, si bien el análisis de una frase, para ser justo y completo, requiere la consideración de su contexto textual, ver intertextual —el texto es así la unidad mínima del análisis.

Ya por su nombre, la textometría afirma también su adhesión a la unidad texto: la evolución en la designación de la “lexicometría” en “textometría” quiere expresar que el análisis realizado no se limita al estudio del léxico, sino que se interesa, antes que nada, en la descripción del texto en sus múltiples dimensiones. Desde su origen, la textometría trabaja naturalmente sobre *corpus* de textos integrales, por oposición a las prácticas de *corpus* de frases o de fragmentos de textos. De hecho, la textometría es empleada tanto por lingüistas como por investigadores de otras ciencias humanas (literatura, política, historia, filología, etc.), cuyo objetivo es el de darse los medios de renovar la lectura de sus *corpus*, respetando los datos obtenidos.

El textómetro conoce muy bien su *corpus*, cuando no, lo adquiere. Ya lo ha recorrido, en ciertos casos puede ya incluso haberlo leído y releído, y —¿fuente o resultado de esta frecuenciación asidua?— está a menudo unido a él. Ya que el acercamiento textométrico es aquel de la curiosidad de una lectura renovada mediante la puesta en evidencia de regularidades todavía

no percibidas.²⁶ La textometría es así complementaria de acercamientos como la búsqueda o el filtraje de informaciones, como los sistemas de pregunta-respuesta, donde el *corpus* es de alguna manera un reservorio, a veces constituido sobre la marcha por filtraje sobre algunos criterios como la presencia de ciertas palabras-clave, *corpus* reservorio cuyo conocimiento global importa poco, y cuyo papel es sólo el de poder entregar algunos extractos al contenido “pertinente”, a menudo sin consideración particular por la formulación empleada y su contextualización completa. La textometría se caracteriza aquí de manera muy clara por su adhesión a los textos que componen su *corpus*, construido y estudiado por sí mismo. En este sentido, Geffroy y Lafon (1982) habían tenido que subrayar, no sin humor, “l’insécurité dans les grands ensembles”²⁷: la aplicación de cálculos textométricos a los *corpus* demasiado grandes para tener un primer conocimiento interior, no superficial, como a los *corpus* donde los textos están fundidos en algunas grandes categorías meta-textuales que ocultan la unidad de cada texto, apenas proporcionan análisis significativos, pues la interpretación de los resultados de los cálculos es tan solo aproximativa y limitada, e incluso puede extraviar.

La importancia de los textos se materializa en la concepción misma de los programas textométricos. En *Hyperbase*, la hipertextualidad, masivamente empleada, es sistemáticamente aprovechada para regresar al texto y visualizar las ocurrencias

²⁶ De ahí las connivencias notables con los estudios literarios (trabajos sobre las *Flores del Mal* (Baudelaire) de Viprey, sobre el teatro de Giraudoux de Brunet, etc.), la filología (p. ej. interface de consulta de la Base de Francés Medieval) o la exégesis (*cfr.* el interés del Centro Informático y Biblia de la abadía de Maredsous por un programa como *Hyperbase*).

²⁷ Esta expresión es un juego de palabras en francés, ya que se presta a una doble lectura. *A priori*, podría tratarse de un tema debatido en la prensa, significando la violencia y la inseguridad en las zonas urbanas constituidas de inmuebles; pero aquí, en el contexto de estadísticas sobre las palabras, es reinterpretada para evocar los riesgos de error científico en el trabajo sobre los *corpus* demasiado grandes, del cual no se llega a conocer el contenido.

en su contexto textual. La ergonomía de los programas prevé siempre o una visualización de texto a los lados de la visualización de listas, de cuadros o de representaciones gráficas, o una navegación hipertextual, permitiendo un acceso inmediato a los contextos de ocurrencia elegidos.²⁸

3.2.2. La contextualización como principio de análisis y el papel determinante del *corpus* de referencia

La textometría cuenta, sitúa, caracteriza las unidades en los contextos: así se detectan los lazos (morfológicos, lexicales, sintácticos, semánticos...) entre unidades; se establecen igualmente similitudes entre contextos (típicamente entre textos), que producen cartografías dibujando tipologías. Como se ha visto (en términos de modelización), los contextos son tanto locales (coocurrencias, concordancias) como globales (especificidades, cartografía por análisis factorial de correspondencias).

“Para la problemática hermenéutica, [el texto] es la unidad mínima (aunque no elemental). Un texto sólo puede leerse en un *corpus*” (Rastier, 2008). “La comprensión del texto [...] procede por contextualización e intertextualización” (Rastier, 2001: III.3.2). Esto se extiende a todos los planos, reformulándose en varios principios (Rastier, 2001: III.3.2): El principio de contextualidad (“dos signos o dos pasajes de un mismo texto, colocados lado a lado, seleccionan recíprocamente los elementos de significación (semas”), el principio de intertextualidad (análogo para dos pasajes de textos diferentes), y el principio de architextualidad, según el cual todo texto colocado en un *corpus* recibe determinaciones semánticas y modifica potencialmente el sentido de cada uno de los textos que lo componen. Proporcionando una lectura no lineal (mediante la extracción de contextos, la generación de concordancias), la textometría juega fundamentalmente sobre los principios de contextualidad

²⁸ Ver, por ejemplo, Heiden (2004) para el programa Weblex.

y de intertextualidad. De igual modo, las coocurrencias se han podido comprender como una forma de contextualización mínima en el espíritu de la semántica interpretativa (Mayaffre, 2008). En cuanto al principio de architextualidad, el funcionamiento estadístico del *corpus* de referencia es una concretización: en efecto, todo texto incluido en el *corpus* aporta su contribución a las frecuencias globales,²⁹ y recíprocamente se encuentra caracterizado en relación con las frecuencias globales. La elección del *corpus* de referencia es determinante para el análisis y condiciona completamente la interpretación. La textometría permite ver un texto de manera completamente diferente según el *corpus* sobre el cual es perfilado. El texto no tiene, pues, un contenido por extraer, una sola “buena” lectura sino tantos enfoques posibles como contextualizaciones pertinentes hay en el *corpus*. Por la vía del *corpus* de referencia, lo global determina lo local, y el análisis es la puesta en evidencia de formas que se destacan sobre un fondo (Rastier, 2001: I.4.2.).

3.3. Semántica interpretativa

A un acercamiento ontológico, emparentado con una forma de contemplación (de lo que “es”), la semántica interpretativa opone una concepción dinámica del sentido, una praxeología, relativa a las prácticas contextualizadas. El sentido se construye en el transcurso de la lectura, familiarizada con un reconocimiento de formas, poco a poco delimitadas, e incluso evolutivas. Se trata de una actividad perceptiva, el gesto interpretativo se ajusta en función de limitaciones lingüísticas recibidas del texto. Así se

²⁹ También se puede trabajar con un *corpus* de referencia que no contiene el *corpus* de trabajo (*cfr.* Las especificidades exógenas en *Hyperbase*, calculadas en relación al *corpus* literario Frantext), pero es un caso particular más raro (casi no implementado por los programas de textometría) y a menudo más discutible (estatus y calidad del *corpus* de referencia, conocimiento preciso de su composición y acceso a los textos, adecuación y posibilidades de ajustamiento al *corpus* de trabajo).

concibe la riqueza de sentido del texto, pero también su significatividad, no arbitraria, las limitaciones lingüísticas impidiendo hacer decir cualquier cosa a un texto.

El análisis textométrico procede igualmente de un enfoque construido: no se puede proporcionar un *corpus*, “poner en marcha” el programa, y recuperar el resultado como producto terminado. Por el hecho de su importancia determinante, la constitución del *corpus* es una primera etapa que compromete elecciones interpretativas: los datos no están precisamente dados (Rastier, 2008). La codificación de ese *corpus*, y su puesta en correspondencia con la estructura de datos textométricos (para definir las unidades, las tipologías, las divisiones, etc.) proceden igualmente de elección, en correspondencia con las hipótesis y expectativas interpretativas. Luego, se tratará de hallar un buen punto de entrada; de lanzar un cálculo pertinente, comprendiendo los principios según los cuales funciona; de calificar los resultados cuantitativos, con un flujo que siempre remite a tareas de lectura, de comprensión (recorrido de contextos, comparación, reagrupamientos). La dinámica de la interpretación se traduce una vez más por la elección de un nuevo punto de entrada, de un nuevo contexto, o de un nuevo cálculo, que a menudo es de hecho un ajuste del tratamiento precedente —y encontramos muy concretamente la interpretación como acción y como gesto que afina dinámicamente su trayectoria.

Podemos reconocer en ciertos procesos textométricos principios hermenéuticos tradicionales, como el de los “pasajes paralelos”, que consiste en recurrir, para la comprensión de un pasaje oscuro, a la lectura de otro pasaje del mismo texto (o de un texto relacionado) que aborda el mismo tema. La *delinearización* y las reorganizaciones del texto facilitadas por la digitalización equipan eficazmente esta técnica hermenéutica de pasajes paralelos (Bourion, 2001: 116; Pincemin, 2006).

3.4. Semántica diferencial

La semántica interpretativa se define como una semántica diferencial, por oposición a las semánticas inferenciales o referenciales. Es, por esa razón, plenamente lingüística pues no está fundada en una realidad exterior, física o psíquica —incluso si permite comprender posteriormente la articulación del texto con esas realidades de otro orden.³⁰

3.4.1. El funcionamiento diferencial de las representaciones y los cálculos textométricos

Formalmente, en el plano de los descuentos de frecuencias y otros cálculos que de ahí resultan, la textometría supone que de acuerdo a una tipificación de unidades, que fija lo que se recuperó del mismo (y se acumula en términos de frecuencias) y lo que es diferente (y participa, por lo tanto, en el número de tipos), ya tenemos fundamentalmente una modelización que, en un momento dado, elige asimilar ciertas unidades y disimilar otras. La tipificación, que regula ese juego de identificación y de oposición, es relativa al punto de vista elegido, que puede variar en el transcurso del análisis —al igual que, lingüísticamente, las aproximaciones y las diferenciaciones evolucionan según la activación, la aferencia o la inhibición de semas, dependientes ellos mismos en particular de las contextualizaciones.

Luego, las estadísticas contrastivas propuestas en textometría han sido bien comprendidas como un mecanismo diferencial. Debido a que estas estadísticas ponen en evidencia lo que se aleja de la norma (definida por el *corpus* de referencia), ellas destacan los contrastes en un todo que hace sistema (Bourion, 2001: 18; Rastier, 2001: III.2.2).

³⁰ Aquí se reenvía al concepto de *polos intrínsecos* del texto y su participación en la definición de géneros textuales (Rastier, 2001: I.1.1).

Pero también, siempre en los procedimientos textométricos, y sin ir hasta procedimientos matemáticos elaborados, un simple orden alfabético de una lista del vocabulario, o un orden del contexto izquierdo o derecho de una concordancia, hacen también entrar en juego heurísticamente en su lectura una percepción diferencial: el ojo destaca los motivos recurrentes, acercados por el ordenamiento, y las variaciones en el interior o en las fronteras de estos reagrupamientos (Pincemin, 2006).

3.4.2. La atención en la estructura y la dominación de lo cualitativo

La semántica interpretativa resulta de un acercamiento estructural de la lengua, en la línea de los trabajos de Hjelmslev, Greimas, Pottier y Coseriu. Se trata de situar las unidades lingüísticas unas en relación con otras, en el interior del sistema que forman, y no por una cualificación o una evaluación que reenvía a una realidad externa a la lengua. Esto vale también para la atribución de una interpretación: la semántica interpretativa no pretende asociar a un signo, una proposición o un texto, su interpretación, sino que busca las restricciones planteadas por la lengua y por su uso en el texto considerado para jerarquizar las lecturas posibles.

Asimismo, ciertas medidas textométricas corresponden a probabilidades o a valores estadísticos que permiten una cualificación en términos de significatividad estadística; otras son simples descuentos de frecuencias, con ciertos valores particulares (como la frecuencia 1 del hápax); estas cifras son utilizadas como umbral de los resultados y para la determinación de una selección (de palabras, de coocurrencias, etc.). Pero también, y sobre todo, son utilizadas para la clasificación que inducen, el “ordenamiento jerárquico” que permiten operar. La práctica textométrica consiste, entonces, luego de la generación de una lista delimitada y ordenada por criterios cuantitativos, en trabajar más cualitativamente sobre zonas de la lista: el inicio de la

lista que proporciona los elementos dominantes, como zonas intermedias (para observar los fenómenos más matizados), incluso cercanos del umbral (para ajustar la selección). Dicho de otra manera, lo cuantitativo guía el análisis, pero es un examen cualitativo (con retorno al texto, etc.) el que determina la interpretación.

3.5. Semántica unificada

La semántica interpretativa está unificada, en el sentido en que los principios diferenciales e interpretativos se observan en diferentes planos, típicamente aquellos de la lexía, del período y del texto. Quizás también podríamos situar bajo esta preocupación unificadora el rechazo de las divisiones disciplinarias que separan puntos de vista no obstante complementarios e íntimamente articulados: sintaxis, semántica y pragmática —una semántica bien hecha integra consideraciones sintácticas y debe saber describir los fenómenos relegados a la pragmática—, filología y hermenéutica, y más generalmente las “artes y ciencias del texto” sobre las cuales Rastier (2001) propone adoptar un punto de vista que abarca, por el hecho de su objeto en común, el texto.

La textometría se adapta muy naturalmente con la multiplicidad de planos y sus analogías de funcionamiento. De hecho, formalmente, la técnica no exige más que proporcionarle unidades o “contenidos”, repartidos (mejor, contextualizados) en los “continentes”. Poco le importa la naturaleza de esos contenidos y continentes, al lingüista le corresponde determinar las unidades pertinentes, y de hacerlas variar como bien le parezca. Las unidades no son necesariamente las palabras; las entidades englobantes, no necesariamente los textos.

Como lo hemos visto, las unidades utilizadas por el cálculo no tienen pretensión lingüística, las unidades lingüísticas/hermenéuticas son construidas, y lo que queremos señalar aquí es que esas unidades construidas pueden relevar de otros planos



que el de las unidades que hayan servido al cálculo. Claramente, la textometría se sitúa así del lado de los acercamientos textuales, pues ella no es simple malabarismo con signos predefinidos. Como lo muestran las experiencias de Brunet (2006b),³¹ si hay signo no es considerado en tanto tal, sino que es aprehendido al vuelo para un objetivo más global, el de hacer destacar líneas de fuerzas; es decir, despejar las formas significantes. Un cálculo basado en palabras puede así conducir a localizar semas (infralexicales, en el sentido en que una palabra es *a priori* portadora de varios semas) y a construir moléculas sémicas que representan un tema (supra-lexical, en el sentido en que su manifestación puede ser difusa sobre todo un pasaje, ver todo un texto, y donde se presta a lexicalizaciones múltiples). Rastier hace hincapié en cómo, incluso trabajando en apariencia en el nivel lexical del *corpus* se captan de hecho, realidades de otro nivel:

Para progresar, la temática debe rebasar el análisis léxico [...]. La palabra a partir de la cual puede comenzar la investigación temática no es el objeto, a diferencia de una palabra-vedette que sería objeto de una investigación lexicográfica. Desde luego se va a buscar, utilizando los medios informatizados de asistencia, otras palabras y expresiones que sean co-ocurrentes. Una vez interpretadas, las co-ocurrencias para las que se habrá identificado una relación semántica serán considerados como *correlatos*, es decir, como lexicalizaciones complementarias de la misma molécula sémica. La red de correlatos conecta las manifestaciones lexicales del tema. Pero se requiere poder discernir los mejores puntos de entrada en esa red: La ‘vedette’ es sólo uno de esos puntos de entrada que presume lexicalizar sintéticamente el tema que se trata de describir Rastier (2001: VII.3.2).

Las experiencias de caracterización de textos o de géneros, basadas explícitamente en unidades no lexicales (medidas morfosintácticas del *corpus* Synapse, índices semióticos del

³¹ Y la de otros antes que él; Salem, por ejemplo, había tenido experiencias similares.

corpus Princip, etc.) han mostrado que las técnicas estadísticas permiten captar regularidades significativas, aunque imperceptibles para una lectura sin la herramienta textométrica. La textometría entonces formaría parte de las técnicas capaces de sumergirse en las dimensiones profundas del material textual:

Por último, la oposición humboldtiana entre la forma interior y la forma exterior de los textos, que ha hecho correr tanta tinta entre los estilistas, podría recibir una nueva formulación que la relativiza: la forma interior, lejos de ser un misterio estético, está constituida por las regularidades hasta el presente imperceptibles de la forma exterior, la de la expresión, que los medios teóricos y técnicas de la lingüística de *corpus* permiten evidenciar en el presente. En otros términos, el contenido de un texto no se reduce, desde luego, a una misteriosa representación mental: un texto está hecho de dos planos, el de las formas semánticas y el de las formas expresivas, cuyo género, en particular, norma la puesta en correlación. En el seno de cada plano se establecen relaciones forma/fondo de tipo gestáltico, que permiten la percepción semántica y fonológica (Rastier, 2005).

La distinción de dos planos reafirma que la textometría puede hacer percibir las regularidades expresivas que participan en la construcción de formas semánticas, sin que por ello entregue directamente un sentido, “extraer el contenido”. La interpretación permanece como parte integrante del análisis textométrico.

4. Perspectivas de aportes mutuos

Actualmente la textometría suscita nuevas proposiciones y nuevos desarrollos, particularmente en torno de la realización colaboradora de una plataforma informática abierta, agrupando las investigaciones y los desarrollos informáticos de los principales equipos del dominio.³²

³² Este proyecto llamado *Textométrie* (<http://textometrie.ens-lsh.fr/>) es actualmente (2007-2010) financiado por la Agence Nationale de la Recherche (ANR-06-CORP-029).

4.1. Modelización de los textos y de los corpus

El punto de vista de la semántica interpretativa anima a preservar y desarrollar la utilización de los *corpus* estructurados y etiquetados para la pluralidad de segmentaciones y de descripciones que ellos propician (Loiseau, 2006). En el mismo espíritu, la logometría (Mayaffre, 2005), otra designación reciente de la lexicometría (al igual que la textometría), afirma la pertinencia de una textometría capaz de trabajar sobre múltiples niveles lingüísticos, y la semántica interpretativa forma parte de sus fundamentos lingüísticos fuertes.

La semántica interpretativa insiste también en la posibilidad de definir los *sub-corpus* “de pertinencia Enriquecida”. Ella coincide con el retorno de los usuarios de programas de textometría, que constatan la necesidad de poder ajustar y redefinir un *sub-corpus* en el transcurso de los análisis. Esta dinámica del *corpus*, en parte ya asumida en el programa Lexico 3, está inscrita en el cuaderno de especificaciones técnicas de las nuevas aplicaciones textométricas.

La discusión está comprometida con la redefinición de una modelización del texto. El modelo textométrico tradicional está basado en una segmentación de referencia sin necesariamente una valoración teórica asociada pero, a pesar de todo, en relación con la cual todas las otras segmentaciones deben definirse. El punto de vista de la semántica interpretativa sostiene la proposición de un modelo diferencial, en el cual el texto se despliega en el cruce de las dimensiones descriptivas adoptadas, afirmando así la apertura de la modelización (jamás completa), y su carácter diferencial y unificado (que prescinde de introducir jerarquías entre las dimensiones descriptivas) (Pincemin, 2008).

4.2. Concepción de funcionalidades

La semántica interpretativa puede proponer unas pistas para comprender las funcionalidades textométricas existentes indi-

vidualmente (p. ej. el análisis de concordancias, en Pincemin, 2006). Puede renovar también la manera de considerarlas en su conjunto: por ejemplo la concepción morfosemántica, que se apoya en la oposición entre puntos regulares y puntos singulares (Rastier, 2001: I.4.2) podría sugerir un análisis global de las funcionalidades textométricas entre aquellas que captan regularidades, particularmente las repeticiones de lo idéntico, y aquellas que delimitan singularidades.

Las proposiciones de Bourion (2001), concretizadas en los programas realizados por Maucourt, han participado en el mejoramiento de ciertas funcionalidades (cruzamiento de concordantes con cocurrencias) y en la formulación de innovaciones, como los cuadros sinópticos (que quedan por integrar en los programas actuales). Las investigaciones de Malrieu (2006) exploran la actualización del juego de los indicadores adaptados a los textos y los cálculos apropiados a la descripción de estructuras intratextuales. Loiseau (2007) desarrolla en los planos teórico e informático la explotación de los *corpus* multi-anotados, para los análisis que tomen en cuenta las unidades lingüísticas de importancia y de diversas naturalezas, según la idea rastieriana de semántica unificada, con múltiples posibilidades de contextualización. Su programa Corpus Reader opera por el momento unos cálculos no específicamente textométricos, pero explora una vía en la cual la textometría podría extenderse.

Por último, es probable que el terreno de la anotación dinámica del *corpus*, es decir el enriquecimiento del *corpus* por las interpretaciones que pueden servir de apoyo a los análisis ulteriores, pueda interesar a la semántica interpretativa. Una colaboración sobre este aspecto sería mucho mejor recibida que la actualización de una tal funcionalidad, y hacer que los usos asociados condicione aún completamente su pertinencia (mal definida o mal aplicada, una funcionalidad tal puede volver totalmente inmanejable y no interpretable el *corpus*).

4.3. Elaboración de orientaciones metodológica

El trabajo en torno del análisis temático (Rastier, 1995) ha permitido, lo hemos visto, la elaboración de un método de búsqueda de correlatos para la construcción de moléculas sémicas representativas de temas. Las tesis de Bourion (2001) y Deza (1999) han proseguido y precisado esta reflexión metodológica sobre el acceso semántico a los bancos textuales. Las de Poudat (2006) y de Loiseau (2006) las han extendido tomando en cuenta las posibilidades de enriquecimiento lingüístico de los *corpus*, y considerando más sistemáticamente los cuatro componentes de la descripción textual según la semántica interpretativa: temática, dialéctica, dialógica y tác-tica. Si continuaran con el desarrollo de sus experiencias y de sus prácticas en los procedimientos textométricos, los lingüistas de inspiración rastieriana podrían contribuir con la elaboración un conjunto que consistiría de elementos metodológicos³³ fundados en un marco teórico lingüístico fuerte.

Todas las etapas del recorrido textométrico pueden ser clarificadas por una puesta en perspectiva teórica y metodológica: las consideraciones filológicas relacionadas con la constitución de un *corpus* inicial y su codificación, la interpretación semántica de las funcionalidades (concordancias, coocurrencias, *cfr. supra*), los métodos de examen (como la organización de los contextos de concordancia en función de los semas actualizados), el encadenamiento de tratamientos. Estos aportes metodológicos pueden ser asociados con la actualización de las interfaces.

³³ Por ejemplo, la observación de ritmos semánticos sugerida por Bourion (2001, vol. I: 47 y vol. II: 18-19).

Conclusión: perspectivas prácticas y teóricas

Otras técnicas, otros cálculos, podrían ciertamente mostrar la pertinencia para el método de la semántica interpretativa: si la argumentación de este artículo es entusiasta, no quiere pretender, sin embargo, ni la exclusividad de la textometría como propuesta de respuesta informática a la teoría rastieriana³⁴ ni al carácter ideal de la textometría actual, que está de hecho todavía en plena evolución.

Siendo así, es sorprende encontrar tantas connivencias de fondo entre los dos acercamientos, en torno a los fundamentos: texto, contextualización, intertextualidad y *corpus*, semántica diferencial, interpretación y dinámica de la construcción del sentido. Aunque hay mucho interés en proseguir y profundizar las colaboraciones.

La textometría podría ser la base de un entorno de lectura “SAAS” (Sistema de Ayuda al Análisis Semántico), para retomar los términos de Bourion (2001). El interés radica en aprovechar las posibilidades de lo digital para dotarse de herramientas de programación que renueven los recorridos de lectura y los puntos de apoyo interpretativos. En este sentido, el estudio del potencial de la textometría para acompañar un acercamiento rastieriano de los textos releva plenamente de una reflexión sobre semántica e interpretación: ella lo explora en un plano concreto, preciso y revelador. La experimentación de esas nue-

³⁴ Encontraremos unos ejemplos diversificados de perspectivas y de aplicaciones de software en relación con la semántica interpretativa en el sitio *Texto!* (<http://www.revue-texto.net/>), especialmente en las secciones *Dits et inédits* y *Corpus et trucs*. La mayoría tiene relación con la textometría, por su utilización de estadísticas textuales (Rossignol, Mauceri, Reutenaer...) o por el lugar central dado al análisis cualitativo, semi-automatizado, típicamente el de aplicaciones de anotación y de visualización de semas y de recorridos (Beust, Tanguy, Thlivitis, Bénel, Perlerin, Roy...). Pero son también posibles otras vías menos cercanas, como el conexionismo o la programación lógica con restricciones que han retenido la atención por su afinidad con el carácter perceptivo de la semántica (Prié, 1995; Rastier *et al.*, 1994).

vas formas de lectura y de interpretación no es una simple aplicación, que es además inevitablemente reductora, de la reflexión teórica semántica: está en condición de relanzarla y de renovarla, develándole realidades textuales o hermenéuticas aún no percibidas u olvidadas.

Referencias

- ABALALI, Driss; Poudat, Céline (15-19 juin 2009). « Sémantique de corpus. Concepts fondamentaux et dialogue avec l'ADT », *Ecole thématique CNRS Méthodes Informatiques et Statistiques en Analyse de Textes*, Besançon.
- BEAUDOUIN, Valérie (2002). *Mètre et rythme du vers classique - Corneille et Racine-*, Paris : Champion, coll. « Lettres numériques », 2.
- BEAUVISAGE, Thomas (2000). *Exploiter des données morpho-syntactiques pour l'étude statistique des genres. Application au roman policier*, Mémoire de DESS, Centre de Recherche en Ingénierie Multilingue, INaLCO, Paris.
- BOMMIER-PINCEMIN, Bénédicte (6 avril 1999). *Diffusion ciblée automatique d'informations : conception et mise en œuvre d'une linguistique textuelle pour la caractérisation des destinataires et des documents*, Thèse de Doctorat, Linguistique, Université Paris IV (Sorbonne).
- BOURION, Evelyne (1995). « Le réseau associatif de la peur », in François Rastier (éd.), *L'analyse thématique des données textuelles : l'exemple des sentiments*, Paris : Didier, collection Études de sémantique lexicale, 107-145 [<http://www.revue-texto.net/1996-2007/Parutions/Analyse-thematique/Bourion.pdf>].
- _____ (soutenue le 14 décembre 2001). *L'aide à l'interprétation des textes électroniques*, Thèse de doctorat en Sciences du langage, Université de Nancy II [http://www.revue-texto.net/Corpus/Publications/Bourion/Bourion_Aide.html].

- BRUNET, Etienne (2006a). *Hyperbase, logiciel documentaire et statistique pour la création et l'exploitation de bases hypertextuelles. Manuel de référence. Version 6.0 (mai 2006)*. Laboratoire Bases, Corpus et Langage, UFR Lettres, Université de Nice.
- _____ (juin 2006b). « Le corpus conçu comme une boule », *Corpus en Lettres et Sciences sociales : des documents numériques à l'interprétation, Actes du XVIIe Colloque d'Albi Langages et Signification*, Albi, 10-14 juillet 2006, Carine Duteil-Mougel & Baptiste Foulquié (éds), et Texto!, vol. XI, n. 2 [<http://www.revue-texto.net/1996-2007/Parutions/Livres-E/Albi-2006/Brunet.pdf>].
- ERLICH, David (1995). « Une méthode d'analyse thématique. Exemples de l'ennui et de l'ambition », in François Rastier (éd.), *L'analyse thématique des données textuelles : l'exemple des sentiments*, Paris : Didier, collection Études de sémantique lexicale, 85-103 [<http://www.revue-texto.net/1996-2007/Parutions/Analyse-thematique/Erlich.pdf>].
- GEOFFROY, Annie; Lafon, Pierre et Tournier, Maurice (1974). « L'indexation minimale - Plaidoyer pour une non-lemmatisation », E.N.S. de Saint-Cloud, 30 pages - Communication au Colloque sur l'*Analyse des corpus linguistiques : Problèmes et méthodes de l'indexation maximale*, Strasbourg, 21-23 mai 1973.
- _____ (1982). « L'insécurité dans les grands ensembles. Aperçu critique sur *Le vocabulaire français de 1789 à nos jours* d'Etienne Brunet », *MOTS*, 5.
- HEIDEN, Serge (10-12 mars 2004). « Interface hypertextuelle à un espace de cooccurrences : implémentation dans Weblex », *Actes des 7es Journées internationales d'analyse statistique des données textuelles (JADT 2004)*, Gérald Purnelle & al. (éds), Presse universitaires de Louvain, Louvain-la-Neuve (Belgium) [http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2004/pdf/JADT_055.pdf].
- LAFON, Pierre (1980). « Sur la variabilité de la fréquence des formes dans un corpus », *MOTS*, 1, 127-165.

- _____ ; Salem, André (1983). « L'inventaire des segments répétés d'un texte », *MOTS*, 6.
- LAMALLE, Cédric; Salem, André (13-15 mars 2002). « Types généralisés et topographie textuelle dans l'analyse quantitative des corpus textuels », *Actes des 6es Journées internationales d'Analyse statistique des Données Textuelles (JADT 2002)*, Saint-Malo, Annie Morin & Pascale Sébillot (éds), Rennes : IRISA [http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2002/PDF-2002/lamalle_salem.pdf].
- LEBART, Ludovic et Salem, André (1994). *Statistique textuelle* : Dunod.
- LEBART, Ludovic; Salem, André et Bécue, Mónica (2000). *Análisis estadístico de textos*, Lleida : Editorial Milenio.
- LOISEAU, Sylvain (1 décembre 2006). *Sémantique du discours philosophique : du corpus aux normes. Autour de G. Deleuze et des années 60*, Thèse de doctorat, Sciences du langage, Université de Paris X Nanterre.
- _____ (2007). « *CorpusReader* : un dispositif de codage pour articuler une pluralité d'interprétations », *Corpus*, 6 [<http://corpus.revues.org/index1282.html>].
- MALRIEU, Denise (2006). « Familles narratologiques et balisage du roman contemporain », *Proceedings of the First International Conference of the Alliance of Digital Humanities Organisations*, Paris: Centre Cultures Anglophones et Technologies de l'information, Paris IV.
- _____ et Rastier, François (2001). « Genres et variations morphosyntaxiques », *Traitements automatiques des langues*, 42 (2).
- MAYAFFRE, Damon (2005). « De la lexicométrie à la logométrie », *L'Astrolabe* [<http://www.uottawa.ca/academic/arts/astrolabe/articles/art0048/Logometrie.htm>].
- _____ (2008). « De l'occurrence à l'isotopie. Les co-occurrences en lexicométrie », *Syntaxe & Sémantique*, 9.

MELLET, Sylvie; Salem, André (éds) (2009). Topographie et topologie textuelles, *Lexicometrica* [<http://www.cavi.univ-paris3.fr/lexicometrica/numspciaux/special9.htm>].

MÉZAILLE, Thierry (1995). « La couleur des sentiments chez Proust », in François Rastier (éd.), *L'analyse thématique des données textuelles : l'exemple des sentiments*, Paris : Didier, collection Études de sémantique lexicale [<http://www.revue-texto.net/1996-2007/Parutions/Analyse-thematique/Mezaille.pdf>].

PINCEMIN, Bénédicte (2002). « Sémantique interprétative et analyses automatiques de textes : que deviennent les sèmes ? », Benoît Habert (dir.), Dépasser les sens iniques dans l'accès automatisé aux textes, *Sémiotiques*, 17, décembre 1999.

_____ (juin 2006). « Concordances et concordanciers -De l'art du bon KWAC », *Corpus en Lettres et Sciences sociales : des documents numériques à l'interprétation, Actes du XVIIe Colloque d'Albi Langages et Signification*, Albi, 10-14 juillet 2006, Carine Duteil-Mougel & Baptiste Foulquié (éds) et Texto!, vol. XI, n. 2 [<http://www.revue-texto.net/1996-2007/Parutions/Livres-E/Albi-2006/pincemin.pdf>].

_____ (2008). « Modélisation textométrique des textes », *Actes des 9es Journées internationales d'Analyse statistique des Données Textuelles (JADT 2008)*, Lyon, 12-14 mars 2008, Serge Heiden & Bénédicte Pincemin (éds), Lyon : Presses Universitaires de Lyon, vol. II [<http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2008/pdf/pincemin.pdf>]

POUDAT, Céline (20 juin 2006). *Etude contrastive de l'article scientifique de revue linguistique dans une perspective d'analyse des genres*, Thèse de doctorat, Sciences du langage, Université d'Orléans [<http://www.revue-texto.net/1996-2007/Corpus/Publications/Poudat/Etude.html>].

PRIÉ, Yannick (1995). *Contribution à une clarification des rapports entre Sémantique Interprétative et Informatique*, Mémoire de DEA, Informatique, Université de Rennes 1 [<http://www.revue-texto.net/1996-2007/Inedits/Prie95.pdf>].

- RASTIER, François (1987). *Sémantique interprétative*, Presses Universitaires de France. Traducción de Eduardo Molina y Vedia: *Semántica interpretativa*, México: Siglo XXI, 2005.
- _____ (1991). *Sémantique et recherches cognitives*, Presses Universitaires de France.
- _____ (éd.) (1995). *L'analyse thématique des données textuelles : l'exemple des sentiments*, Paris : Didier, collection Études de sémantique lexicale.
- _____ (2001). *Arts et sciences du texte*, Presses Universitaires de France. Traducción de Enrique Ballón Aguirre: *Artes y ciencias del texto* [en prensa].
- _____ (2005). « Enjeux épistémologiques de la linguistique de corpus », in Geoffrey Williams (éd.), *La Linguistique de corpus*, Rennes : Presses Universitaires de Rennes [http://www.revue-texto.net/1996-2007/Inedits/Rastier/Rastier_Enjeux.html].
- _____ (12-14 mars 2008). « Que cachent les «données textuelles» ? », *Actes des 9es Journées internationales d'Analyse statistique des Données Textuelles (JADT 2008)*, Serge Heiden & Bénédicte Pincemin (éds), Lyon : Presses Universitaires de Lyon, vol. I [<http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2008/pdf/rastier.pdf>].
- _____ ; Cavazza, Marc et Abeillé, Anne (1994). *Sémantique pour l'analyse*, Paris : Masson.
- REUTENAUER, Coralie; Valette, Mathieu et Jacquay, Evelyne (9-11 décembre 2009). « De l'annotation sémique globae à l'interprétation locale : environnement et image sémiques d'«économie réelle» dans un corpus sur la crise financière », *Conférence ARCO « Interprétation et problématiques du sens »*, Rouen.
- SALEM, André et Fleury, Serge (éds.) (2008). « Explorations textométriques », *Lexicometrica* [<http://www.cavi.univ-paris3.fr/lexicometrica/numspeciaux/special8.htm>].

VALCESCHINI-DEZA, Nathalie (29 juin 1999). *Accès sémantique aux bases de données textuelles*, Thèse de doctorat, Linguistique, Université de Nancy 2.

VALETTE, Mathieu (22-25 juin 2004). « Sémantique interprétative appliquée à la détection automatique de documents racistes et xénophobes sur Internet », in Patrice Enjalbert & Mauro Gaio (éds) *Approches Sémantiques du Document Numérique, Actes du 7e Colloque International sur le Document Electronique* et (version légèrement étendue) *Texto!* [http://www.revue-texto.net/1996-2007/Inedits/Valette/Valette_Princip.pdf].

_____ (juin 2006). « Observations sur la nature et la fonction des emprunts conceptuels en sciences du langage », *Corpus en Lettres et Sciences sociales : des documents numériques à l'interprétation, Actes du XVIIe Colloque d'Albi Langages et Signification*, Albi, 10-14 juillet 2006, Carine Duteil-Mougel & Baptiste Foulquié (éds). et *Texto!*, vol. XI, n. 2 [<http://www.revue-texto.net/1996-2007/Parutions/Livres-E/Albi-2006/Valette.pdf>].