

# A FRASEOLOGIA TRANSDISCIPLINAR DOS TEXTOS CIENTÍFICOS: DAS COLOCAÇÕES ÀS ROTINAS RETÓRICAS<sup>1</sup>

Agnès Tutin<sup>2</sup>

Tradução: Anna Carolina Teles<sup>3</sup>

Supervisão e revisão: Patrícia Reuillard<sup>4</sup>

Revisão técnica: Sandra Loguercio<sup>5</sup>

## 1. Introdução

O texto científico tem a ver, sem dúvida, com um socioleto, oriundo de uma comunidade discursiva que compartilha fins comuns (Cf. SWALES, 1990), objetivos retóricos comparáveis e uma sublinguagem específica, caracterizada por traços textuais, pragmáticos, sintáticos e lexicais evidentes (Cf. KOCOUREK, 1991). O artigo científico, por exemplo, sobretudo das ciências duras, é um tipo de texto majoritariamente pré-fabricado, tanto em seu esquema discursivo quanto em suas *expressões fixas* mobilizadas como outros gêneros muito diferentes, a exemplo das cerimônias de casamento, dos anúncios pessoais ou dos processos (GÜLICH e KRAFFT, 1997).

Pode-se considerar que a língua utilizada nesse gênero textual, por vezes chamada de “língua científica geral” (PECMAN, 2004), atravessa, de certa forma, as disciplinas<sup>6</sup> (Cf. a *linguagem acadêmica* de Biber, [2006]). Essa língua científica geral é ilustrada particularmente por um léxico transdisciplinar, diferente da terminologia dos domínios científicos definidos, “um léxico compartilhado pela comunidade científica, aplicado na descrição e na apresentação da atividade científica [...] que não se refere aos objetos científicos das áreas de especialidade, mas ao discurso sobre os objetos e os procedimentos científicos” (TUTIN, 2007a, p.6). Esse léxico transdisciplinar está principalmente na fraseologia e foi objeto de inúmeros trabalhos de pesquisa, sobretudo em inglês (Cf.

---

<sup>1</sup> A presente tradução foi autorizada para ser publicada em português, nos Cadernos de Tradução do IL (UFRGS), pela Editora Presses Universitaires de Rennes (PUR) e pela autora Agnès Tutin. Referência bibliográfica completa do artigo original: TUTIN, Agnès. “La phraséologie transdisciplinaire des écrits scientifiques: des collocations aux routines sémantico-rhétoriques”. In: TUTIN, A.; GROSSMANN, F. (org.) *L'écrit scientifique: du lexique au discours*. Rennes: PUR, p. 27-43, 2013.

<sup>2</sup> Laboratoire de Linguistique et de Didactique du Français Langue Étrangère et Maternelle (LIDILEM), Université Grenoble 3 – Stendhal.

<sup>3</sup> Bacharelada em Letras – Tradução Português/Francês, UFRGS.

<sup>4</sup> Professora do Departamento de Línguas Modernas e do Programa de Pós-Graduação em Letras, Instituto de Letras, UFRGS.

<sup>5</sup> Professora do Departamento de Línguas Modernas e do Programa de Pós-Graduação em Letras, Instituto de Letras, UFRGS.

<sup>6</sup> Embora cada família de disciplinas comporte, sem dúvida, especificidades.

HOWARTH, 1996; GLEDHILL, 2000; PECMAN, 2004<sup>7</sup>. Para o francês, Cf. LOFFLER-LAURIAN, 1995). Essa fraseologia tem um papel fundamental no âmbito desse discurso: ela permite primeiramente ao autor/pesquisador reforçar sua pertença a uma comunidade discursiva com o emprego de uma linguagem codificada (*dans la littérature, nous écartons l'hypothèse* [na literatura, nós descartamos a hipótese<sup>8</sup>]...), e facilita, por seu caráter pré-construído, a escrita.

Por outro lado, a fraseologia recorrente nas diferentes disciplinas é uma porta de entrada privilegiada para o estudo epistemológico dos discursos científicos (cf., por exemplo, certos estudos de caso sobre os verbos no âmbito dos projetos KIAP [FLØTTUM *et al.* 2006a] e Scientext [GROSSMANN *et al.*, 2009; TUTIN, 2010b]). Essa fraseologia de gênero pode, além disso, ser explorada na área das ciências da informação, ora como *corpus* de exclusão nos processos de indexação automática<sup>9</sup>, ora como léxico que permite caracterizar o tipo de texto científico<sup>10</sup>. Por fim, no âmbito didático, a especificidade linguística desses textos foi amplamente estudada, sobretudo no inglês, na perspectiva do ensino acadêmico, *Inglês para Fins Acadêmicos* (Cf., por exemplo, os trabalhos de SWALES, [1990], HYLAND, [2000], COXHEAD, [2002]). No francês, o interesse pelo *Francês para Objetivos Universitários*, mais recente, responde a objetivos similares, visando aos estudantes estrangeiros que precisam se aculturar na língua acadêmica francesa (Cf. PARPETTE e MANGIANTE, 2010; CAVALLA, 2010).

Nosso interesse pelo estudo linguístico do texto científico e as necessidades mencionadas levaram-nos a refletir sobre o funcionamento da fraseologia transdisciplinar nesse gênero textual. Voltamo-nos, em um primeiro momento, para as colocações, que constituem um subconjunto particularmente produtivo da fraseologia científica (Cf. TUTIN, 2007b). Desejamos, em um segundo momento, expandir esse estudo para além da faceta da fraseologia científica transdisciplinar de uma maneira funcional. Nessa perspectiva, a tipologia elaborada deve servir tanto aos estudos linguísticos e ao

---

<sup>7</sup> Destaquemos que a abordagem de Pecman é contrastiva e diz respeito igualmente ao francês.

<sup>8</sup> N.T.: Os exemplos apresentados em francês e em inglês neste texto recebem traduções em português entre colchetes para que o leitor compreenda o sentido, mas, ao mesmo tempo, possa igualmente compreender a estrutura em francês e/ou em inglês.

<sup>9</sup> A indexação automática consiste em extrair os termos chave de um texto especializado. A maioria das técnicas baseia-se em procedimentos estatísticos que extraem o léxico específico do texto com ajuda de medida de especificidade. Ora, o léxico transdisciplinar é um léxico específico do gênero, sendo portanto muito frequente nesse tipo textual, mas não se refere aos conceitos da área. Excluir o léxico transdisciplinar permitirá uma melhor indexação dos textos. Vejamos um exemplo em linguística com a palavra *sujeito* [N.T.: em francês, a palavra remete tanto ao *sujeito gramatical*, quanto ao *objeto* de um estudo.]: *sujeito* é um termo, em um texto de linguística, em contextos como *sujeito de uma frase* ou em associação com unidades lexicais como *frase*, *objeto direto*. De outra parte, em *abordar um 'sujet'* [objeto] ou *o 'sujet'* [objeto] *abordado*, ele é, sem dúvida, um não-termo. Desse modo, considerar o ambiente lexical das palavras (colocações e outras expressões polilexicais) permite afinar o processo de indexação. Esse tema é abordado no projeto ANR-ContintTermith, coordenado pelo ATILF e do qual participa o LILIDEM.

<sup>10</sup> O artigo é uma revisão do estado da arte, um artigo principalmente do tipo teórico ou do tipo experimental? O léxico é uma porta de entrada privilegiada para caracterizar o tipo do artigo.

estabelecimento de padrões ou modelos para as aplicações de processamento automático previstas<sup>11</sup> quanto às aplicações didáticas que queremos desenvolver a longo prazo.

Após apresentar uma tipologia das sequências polilexicais aplicada ao léxico transdisciplinar do texto científico, ocupamo-nos de um tipo de expressão no limite da fraseologia, as rotinas retóricas, de que propomos padrões inspirados nos *frames* semânticos de Fillmore (FILLMORE *et al.*, 2003). Mostramos então como esses *frames* foram implementados em um sistema operacional de processamento de *corpus* no projeto Scientext.

## 2. Uma tipologia das sequências polilexicais do texto científico

Nosso objeto de estudo concerne à fraseologia do texto científico, que compreende, em sentido amplo, as sequências polilexicais recorrentes e específicas ao gênero texto científico, sejam fixas ou não, e seja qual for o estatuto sintático dos elementos constituintes da expressão. Essa fraseologia, própria ao texto científico, não é uma terminologia, na medida em que não remete aos conceitos chaves da área (por exemplo, a Linguística ou a Geologia). Ela passa até certo ponto pelas disciplinas, mas conjuntos lexicais próprios às famílias de disciplinas podem, sem dúvida, ser extraídos. Essa fraseologia metacientífica e metadiscursiva essencialmente abstrata descreve os procedimentos científicos, o raciocínio científico, a filiação científica, o posicionamento, o metatexto e a estrutura do discurso. Ela não está necessariamente ausente da língua comum, mas pode assumir acepções específicas nesse gênero ou apenas ser muito frequente.

Esse léxico foi objeto de inúmeros trabalhos em inglês, sobretudo no âmbito do *Inglês para Fins Acadêmicos*, sejam trabalhos sobre o discurso acadêmico e científico em uma perspectiva discursiva (HYLAND, 2005), sejam em uma perspectiva textométrica (BIBER, 2006). De mesmo modo, desenvolveram-se trabalhos sobre as colocações, muito produtivas nesse gênero, sejam colocações lexicais (HOWARTH, 1996), sejam enquanto um fenômeno de definição mais ampla (GLEDHILL, 2000). Quanto ao francês, os trabalhos sobre essa fraseologia são menos numerosos, com exceção dos de Pecman (2004; 2007), que detalharemos adiante, assim como estudos mais pontuais sobre as colocações (TUTIN, 2007b). Destaca-se que a fraseologia usada nos textos científicos parece extremamente diversa e que o protótipo da sequência polilexical não-composicional no plano semântico, fixa no plano sintático, e cujos referentes não são atualizados, mostra-se muito limitante. Não se trata apenas de identificar desordenadamente *pacotes lexicais* à maneira de Biber (2006), mas também de caracterizá-los nos planos discursivo, enunciativo, semântico e sintático.

A tipologia que propomos é inspirada ao mesmo tempo em Burguer (1988), Granger & Paquot (2008) e Mel'čuk (2011). O parâmetro mais expressivo é o papel funcional que as expressões cumprem no discurso científico. A tipologia proposta é

---

<sup>11</sup> No âmbito do projeto Scientext, mas também no do projeto ANR Termith na indexação automática.

provavelmente transponível a outros tipos de discurso, no entanto as expressões extraídas são, sem dúvida, próprias ao gênero estudado. A tipologia elaborada deve atender a várias exigências, muitas vezes difíceis de conciliar:

- Deve ser evidentemente delimitada por critérios claros e parecer funcional para o estudo linguístico, em especial para o estudo do discurso científico.
- Deve ser suficientemente intuitiva, em particular para o tratamento semântico, para um tratamento didático.
- Deve ser satisfatoriamente precisa para permitir o estabelecimento de padrões em um Processamento Automático da Linguagem (TAL em francês), que possibilite uma extração informatizada a partir de *corpora* textuais.

A tipologia proposta conduz a quatro tipos de expressões que serão detalhadas a seguir:

1. Sequências polilexicais de função referencial, colocações ou expressões fixas, (ex.: *faire une hypothèse* [levantar uma hipótese]; *point de vue* [ponto de vista])
2. Sequências polilexicais de função discursiva (ex.: *en d'autres termes* [em outras palavras], *pour conclure* [para concluir]...)
3. Sequências polilexicais de função interpessoal, essencialmente expressões de função modal (ex.: *Il est probable* [é provável], *contre toute attente* [contrariando as expectativas]...)
4. Rotinas retóricas, próprias ao discurso científico (ex.: *comme on peut voir sur la figure* [como se pode ver na figura], *le tableau X* [a tabela X]).

### 2.1. Sequências polilexicais de função referencial

Essas expressões referem-se a conceitos simples ou complexos no texto científico, evocados pelo locutor, sem função discursiva ou comunicativa específica<sup>12</sup>. No âmbito de um modelo funcional, como a gramática sistêmico-funcional de Halliday (1994), a metafunção será aqui sobretudo *ideacional*, ou seja, para definir essas expressões não é necessário explicar seu contexto de produção ou sua função discursiva. Do ponto de vista da estrutura semântica interna, podem ser frases completas ou locuções completas (MEL'ČUK, 2011), ou seja, expressões como *prendre en compte* [levar em conta], *tenir compte* [ter em conta], *faire partie* [fazer parte], *point de vue* [ponto de vista], que são não-composicionais no plano semântico e muito restritas no plano sintático. Essa classe de expressões inclui, de mesmo modo, as colocações lexicais (expressões recorrentes, essencialmente binárias, cujos elementos apresentam uma relação sintática), como *faire une hypothèse* [levantar uma hipótese], *réfuter une hypothèse* [refutar uma hipótese], *résultats prometteurs* [resultados promissores] etc. (Cf. GROSSMANN e TUTIN, 2003; TUTIN, 2007b). Para ilustrar, apresentamos, no Quadro 1, um subconjunto de colocações transdisciplinares de tipo V-N extraídas de um *corpus* analisado sintaticamente

---

<sup>12</sup> Esta categoria é, de certo modo, definida pela exclusão das outras.

(expressões mais frequentes em três disciplinas científicas: medicina, linguística e economia). Nota-se que muitas dessas expressões não são exclusivas dos textos científicos, mesmo que algumas assumam, nesse contexto, uma acepção específica (por exemplo, *répondre à une question* [responder a uma questão]<sup>13</sup>).

<i>JOUER un rôle</i> [EXERCER um papel]
<i>AVOIR un effet</i> [TER um efeito]
<i>POSER un problème</i> [LEVANTAR um problema]
<i>OBTENIR des résultats</i> [OBTER resultados]
<i>AVOIR un impact</i> [TER um impacto]
<i>PRÉSENTER les résultats</i> [APRESENTAR os resultados]
<i>FAIRE l'hypothèse</i> [LEVANTAR a hipótese]
<i>PRENDRE la forme</i> [ASSUMIR a forma]
<i>ATTEINDRE l'objectif</i> [ATINGIR o objetivo]
<i>AVOIR une valeur</i> [TER um valor]
<i>DONNER des résultats</i> [FORNECER resultados]
<i>RÉPONDRE à une question</i> [RESPONDER a uma questão]
<i>UTILISER une méthode</i> [UTILIZAR um método]
<i>AVOIR une fonction</i> [TER uma função]
<i>AVOIR un rôle</i> [TER um papel]

Quadro 1: Algumas colocações frequentes em francês do tipo V-Nnos textos científicos.

O tratamento semântico e sintático desses dois tipos de expressões deve ser evidentemente diferenciado: assim como Mel'čuk em *Lexicologia Explicativa e Combinatória* (MEL'ČUK, 2011), acreditamos que os frasemas completos, não-composicionais, devem ser tratados como palavras simples. Eles serão, por exemplo, caracterizados com a ajuda de uma etiqueta semântica (*objet scientifique abstrait* [objeto científico abstrato], *verbe d'opinion* [verbo de opinião]), enquanto as colocações, composicionais, podem constar na base. As expressões *faire une hypothèse* [levantar uma hipótese], *réfuter une hypothèse* [refutar uma hipótese], *hypothèse valide* [hipótese válida] podem assim ser codificadas na entrada *hypothèse* [hipótese], já que cada elemento da colocação é objeto de um tratamento semântico (Cf. TUTIN, 2007b).<sup>14</sup> No que concerne à extração automática das expressões, as técnicas divergem também de acordo com o tipo de sequências. As colocações, geralmente binárias, são frequentemente extraídas com a ajuda

<sup>13</sup> Nesse contexto, *question* [questão] é quase sempre sinônimo de *problématique* [problemática], *problème* [problema], mais do que *phrase interrogative* [frase interrogativa].

<sup>14</sup> O tratamento sintático desses elementos é igualmente diferenciado. A maior parte das colocações, enquanto sequências polilexicais composicionais, tem, na verdade, um comportamento sintático *normal* (por exemplo, *faire DET hypothèse* [levantar DET hipótese] se torna relativa e se reduz: *l'hypothèse qu'il a faite* [a hipótese levantada por ele], *l'hypothèse faite ici* [a hipótese elaborada aqui] etc.). Os frasemas completos são, em geral (mas nem sempre), menos variáveis no plano sintático.

de medidas estatísticas de correlação, que selecionam as expressões de duas palavras que mantêm uma relação sintática e cuja coocorrência é a mais significativa (SERETAN, 2011; DIWERSY e KRAIF, 2012). A situação mostra-se mais complicada para os frasemas completos, na medida em que a expressão pode ser composta por mais de dois elementos e corresponder a uma organização sintática menos previsível. Técnicas com base em bancos de árvores e medidas estatísticas<sup>15</sup> apresentam, entretanto, resultados bastante encorajadores (CORMAN, 2012).

## 2.2. Sequências polilexicais de função discursiva (ou marcadores discursivos)

A noção de *marcador discursivo* recebe, de acordo com as escolas de pensamento, contornos bem flutuantes, e convém precisar qual sua acepção aqui. Nós os definimos como elementos linguísticos que servem para elencar segmentos discursivos, orações ou segmentos textuais maiores, para estruturar a argumentação e a organização textual (*dans un premier temps* [em um primeiro momento], *pour conclure* [para concluir], *en ce qui concerne* [no que concerne], *en d'autres termes* [em outras palavras], *par opposition à* [em oposição a], *c'est pourquoi* [é por isso que]...). Os marcadores discursivos não se limitam portanto aos conectores e podem ter também uma função metadiscursiva, por exemplo, uma operação de reformulação, como a desempenhada por *en bref* [em resumo], ou de paralelismo (*d'une part... d'autre part* [por um lado... por outro lado]) (Cf. PÉRY WOODLEY, 2000)<sup>16</sup>. Contrariamente a Siepmann (2007), não integraremos nessa classe as expressões de função interpessoal, como aquelas que exprimem a certeza, a dúvida, a necessidade e cuja função primeira não é articular segmentos textuais (*il ne fait aucun doute* [sem dúvida alguma], *il est probable* [é provável], *il apparaît indispensable* [parece indispensável] etc.), mas sim expressar uma modalização, um ponto de vista.

No texto científico, as sequências polilexicais de função discursiva cumprem várias funções: servem para guiar o locutor ao longo do texto (função metatextual) e para introduzir definições, exemplos, explicações, para expressar as relações lógicas e para estruturar o raciocínio, entre outras. Do ponto de vista sintático, esses elementos são basicamente adverbiais (em sentido amplo), locuções de função prepositiva ou conjuntiva, ou, até mesmo, sequências de base verbal (ex.: *il en découle que* [disso decorre que]). Siepmann (2007) observa acertadamente que essas expressões são muito mal estudadas e recenseadas, sobretudo no que tange aos marcadores de segundo nível, ou seja, os marcadores de frequência média<sup>17</sup>. Evidentemente essas sequências polilexicais de função

---

<sup>15</sup> A medida utilizada é a Informação Mútua.

<sup>16</sup> Seguindo o modelo de Charolles (2003), não integramos às sequências de função discursiva os *cadratifs* [N.T.: *enquadres enunciativos* conforme o uso de Ingedore Koch, 2006.] verídicos (ex.: *l'année dernière* [no ano passado] ou não verídicos (ex.: *à notre grand regret* [para nosso grande pesar]). Os primeiros serão considerados como sequências de função referencial, enquanto os segundos pertencem à categoria das sequências de função interpessoal (cf. seção 2.3).

<sup>17</sup> Os marcadores de primeiro nível seriam, desse modo, expressões como *en effet* [com efeito], *en revanche* [em contrapartida] etc.

discursiva não são específicas para a maior parte dos textos científicos e podem ser encontradas em outros gêneros, como a vulgarização científica ou os textos jornalísticos. Será dada maior atenção às que são muito frequentes no gênero de nosso interesse<sup>18</sup>.

O tratamento linguístico dessas expressões não é trivial. Além de sua descrição semântica e discursiva, que parece complexa por sua ambiguidade, mostra-se essencial indicar seu modo de emprego sintático (por exemplo, a sequência aparece necessariamente no começo da frase? Sobre qual tipo de constituintes ela incide?). A extração dessas expressões, cujos elementos são adjacentes em geral, é mais simples do que a de frases completas. Uma simples extração de segmentos repetidos (Cf. LAFON e SALEM, 1983) a partir de um *corpus* lematizado, ainda que produza bastante ruído, é normalmente suficiente para a localização desses elementos<sup>19</sup>.

### 2.3. Sequências polilexicais de função interpessoal

Essas expressões são indissociáveis da situação de comunicação e remetem à dimensão interpessoal da linguagem. Consideramos, nessa categoria, as sequências polilexicais de função pragmática ou *pragmatemas*<sup>20</sup> (Cf. MEL'ČUK, 2011; ver também os *enunciados ligados* a uma situação de interação, conforme proposto por Fónagy, [1997]), esses enunciados autônomos – frases padrão ou frases sem verbo, por exemplo, *y'a pas de quoi* [não tem de quê]. *Je vous en prie* [Disponha]. *Après vous* [Você primeiro] –, inseparáveis de situações enunciativas específicas, sobretudo em um contexto de diálogo. Essas expressões são relativamente ausentes nos textos científicos fora do peritexto científico<sup>21</sup>, mas aparecem, com mais frequência, em contextos científicos orais<sup>22</sup>. Incluímos ainda, nas sequências polilexicais de função interpessoal, as sequências polilexicais de função modal, que servem para exprimir o ponto de vista do locutor quanto ao conteúdo oracional. Trata-se de expressões como *semble-t-il* [ao que parece]; *de mon point de vue* [de meu ponto de vista]; *selon toute probabilité* [provavelmente]; *il apparaît nécessaire de...* [parece ser necessário...]; *il est probable...* [é provável...], que, em nossa visão, não se classificam como marcadores discursivos na medida em que sua função primeira não é relacionar conteúdos de segmentos textuais, mas, muito mais, exprimir o ponto de vista do locutor sobre o conteúdo oracional apresentado. Pelo distanciamento que tomam do conteúdo oracional, essas sequências podem, de certa forma, ser analisadas como tendo uma função metadiscursiva, no sentido definido por Hyland (2005)<sup>23</sup>. Do

---

<sup>18</sup> Nesse caso, medidas estatísticas, como a especificidade, podem ser utilizadas.

<sup>19</sup> A extração de base sintática tem a vantagem de gerar menos ruído, ou seja, menos respostas inapropriadas.

<sup>20</sup> A noção de pragmatema é mais restrita, visto que ela não considera necessariamente a polilexicalidade.

<sup>21</sup> Agradecimentos redigidos com fórmulas estereotipadas, por exemplo.

<sup>22</sup> Como nas interações orais de palestras. Por exemplo, o mediador após a fala de um palestrante: *any question?* [alguma pergunta?] ou *nous vous remercions encore pour cet exposé X* [agradecemos mais uma vez por essa apresentação X], em que X corresponde a um adjetivo avaliativo de polaridade positiva.

<sup>23</sup> Esse tipo de expressão polilexical assinala principalmente os primeiros e segundos tipos. Segundo Hyland: [Os princípios chave] são: 1. metadiscorso é distinto dos aspectos propositivos do discurso; 2. metadiscorso se

ponto de vista sintático, essas expressões são principalmente adverbiais (ex.: *selon toute apparence* [ao que tudo indica], *à première vue* [à primeira vista]), ou orações principais que introduzem o conteúdo oracional (*il est probable que* [é provável], *il est discutable de* [é discutível que] etc.). Como para as sequências discursivas, a extração dessas expressões, em que os elementos são muitas vezes adjacentes, pode ser realizada com auxílio da localização de segmentos repetidos.

### 3. Rotinas retóricas

Essas expressões se situam no limite da categoria fraseológica, na medida em que correspondem, por um lado, a enunciados completos e, por outro, comportam um léxico e estruturas sintáticas diversificadas. Trata-se de enunciados estereotipados que remetem a funções retóricas específicas ao gênero de nosso interesse, como a filiação científica indicada por um autor, a demarcação relativamente a um colega de área, a avaliação de atividades científicas de pares, a elaboração de provas factuais: por exemplo, *comme on peut le voir ci-dessous* [como se pode ver abaixo], *on l'a vu ci-dessus* [vimos acima]...

Essas rotinas estão mais associadas a uma função retórica (explicar, persuadir, avaliar, provar) do que a uma estrutura discursiva (estruturar o texto, indicar a relação entre os segmentos textuais); em contrapartida, elas põem em jogo a estrutura enunciativa do discurso por meio dos actantes em questão (por exemplo, o *je* [eu] autor do texto ou o *on* [sujeito indeterminado] que pode incluir o leitor testemunha) e utilizam um material lexical específico. Elas foram estudadas e descritas como padrões principalmente para o TAL no âmbito de trabalhos sobre a extração de informação, como veremos a seguir.

Estabelecemos padrões ou modelos – ou seja, modelizamos – através de esquemas abstratos, mais ou menos como os *frames* semânticos de Fillmore (2003). Mas, contrariamente a essa abordagem de inspiração semântica e cognitiva, esses esquemas têm aqui uma função retórica e são especificamente associados a esse gênero. Por exemplo, a frase *nous nous différencions de Schmigh (1997) sur ce point* [nós nos diferenciamos de Schmigh (1997) nesse ponto] pode ser analisada a partir de um esquema complexo em que entram em jogo um agente (*nous*), uma operação de *demarcação* (*différencions*), uma fonte (*Schmigh (1997)*), um objeto (*ce point*). As rotinas retóricas podem sem dúvida integrar sequências polilexicais mais reduzidas, em especial sequências polilexicais referenciais ou modais. Detalharemos esse modelo após esta seção.

Evidentemente essa tipologia de inspiração funcional não está isenta de problemas. Certos elementos podem se encaixar em várias categorias, e uma escolha deverá ser feita, em caso de ambiguidade, para as implementações do modelo. Por exemplo, as colocações que indicam uma avaliação (*modèle adapté* [modelo adaptado], *hypothèse valide* [hipótese válida], *résultat encourageant* [resultado encorajador]) correspondem, ao mesmo tempo, a expressões referenciais (dizem respeito a objetos específicos) e a sequências de função

---

refere a aspectos do texto que incorporam a interação escritor-leitor; 3. metadiscursos se referem apenas a relações que são internas ao discurso. (HYLAND, 2005, p.38).



interpessoal (expressam um ponto de vista). Para a classificação das expressões, privilegiaremos a estrutura sintática da sequência polilexical. Os exemplos mencionados, que têm uma estrutura colocacional, serão assim preferencialmente classificados como expressões referenciais. As etiquetas semânticas associadas às expressões poderão ser aproximadas de sequências polilexicais de função modal (ex.: *de mon point de vue* [de meu ponto de vista]).

O Quadro 2 sintetiza a tipologia proposta.

Tipo	Definição	Subtipo	Definição	Tipos Sintáticos	Exemplos no Texto Científico
Sequências polilexicais de função referencial	Expressões recorrentes que remetem a noções no texto científico	Frasemas completos (ou expressão fixa)	Expressões não-composicionais	Principais categorias sintáticas abertas (adjetivo, nome, verbo principalmente)	<i>Tirer parti</i> [tirar partido] <i>Prendre en compte</i> [levar em conta] <i>Point de vue</i> [ponto de vista]
		Colocações lexicais	Expressões recorrentes, composicionais, geralmente binárias.	Associações sintáticas de tipo <i>Nom-suj-Verbe</i> [N-suj-V]; <i>Verbe-obj-N</i> [V-obj-N.]; <i>Adj-Epit-Nom</i> [Adj-adj-N]; <i>Nom-Prep-Nom</i> [N-Prep-N]	<i>Faire une hypothèse</i> [levantar uma hipótese] <i>Résultats prometteurs</i> [resultados promissores] <i>Hypothèse de départ</i> [hipótese de partida]
Sequências polilexicais de função discursiva	Expressões recorrentes que servem para estruturar o discurso.			Sequências adverbiais, prepositivas, conjuntivas ou construídas em torno de uma base verbal.	<i>Dans un premier temps</i> [em um primeiro momento] <i>Il en va de même</i> [é o mesmo caso de]...

<b>Sequências polilexicais de função interpessoal</b>	Expressões relacionadas à situação de enunciação.	Sequências polilexicais de função pragmática.	Sequências estereotipadas associadas a uma situação de enunciação específica.	Enunciados geralmente autônomos (frases padronizadas ou frases sem verbo) (quase ausentes do discurso científico escrito).	No campo de agradecimentos nos artigos: <i>Un grand merci à mes collègues X et Y pour leur relecture</i> Adj. [muito obrigado aos meus colegas X e Y por sua leitura Adj.]
		Sequências polilexicais de função modal	Introduzem um ponto de vista sobre o conteúdo oracional.	Elementos adverbiais ou que introduzem orações (completiva ou infinitiva).	<i>En quelque sorte</i> [de certa forma] <i>Il est nécessaire de</i> [é necessário]...
<b>Rotinas retóricas</b>	Enunciados típicos do texto científico que correspondem a uma função retórica específica.		Estruturas frásticas às quais se pode associar um esquema abstrato.	<i>Nous reprenons ici le modèle développé par X</i> [retomamos aqui o modelo desenvolvido por X] <i>Nous nous distinguons de X sur ce point</i> [distinguímo-nos de X nesse ponto]	

Quadro 2: Tipologia de sequências polilexicais no texto científico

### 3.1. Modelos de rotinas retóricas<sup>24</sup>

O texto científico apresenta inúmeros níveis de pré-fabricação linguística, não apenas no que concerne às sequências lexicais e à estrutura discursiva, muitas vezes limitada, mas também na formulação de certas funções retóricas. Essas formulações estereotipadas dos textos científicos, que chamamos de *rotinas retóricas*, situam-se no limite da fraseologia: se correspondem, por um lado, a um tipo de *prêt-à-écrire* do gênero científico e dão frequentemente a impressão de *já terem sido lidas*, também correspondem, por outro, a enunciados completos, construídos em torno de um predicado, e evidenciam

<sup>24</sup> Esta seção retoma, em parte, um capítulo de uma obra (CF. TUTIN, 2010B).

um léxico variado. Mais complexas no plano semântico e mais variadas no plano lexical, diferenciam-se das sequências polilexicais anteriores. Eis alguns exemplos dessas rotinas:

- Fornecer uma prova com a ajuda de um fato: *comme nous le voyons sur ce tableau...*[como podemos ver neste quadro...] *nous pouvons observer que...*[podemos observar que...];
- Contraste e comparação: *Contrairement à Parker (1990), nous...* [Contrariamente a Parker (1990), nós...], *Notre étude diffère de Parker (1990)...* [Nosso estudo difere de Parker (1990)...];
- Filiação científica e acadêmica: *À la suite de Parker (1970), nous...* [Dando seguimento a Parker (1970), nós...], *Nous reprenons la définition de Parker (1979)* [Retomamos a definição de Parker (1979)], *Notre modèle reprend les travaux de Parker (2010)* [Nosso modelo retoma os estudos de Parker (2010)];
- Definir uma problemática: *Notre article traite de la phraséologie scientifique* [nosso artigo trata da fraseologia científica]; *l'objet de notre article est la phraséologie scientifique* [o objeto de nosso artigo é a fraseologia científica].

### 3.2. Rotinas retóricas e modelos da Linguística e do Processamento Automático de Linguagem

As rotinas apresentam uma grande variedade lexical e sintática, no entanto, correspondem a um esquema semântico e retórico recorrente em um tipo de texto. Elas foram, até onde sabemos, bem pouco estudadas pela Linguística, porém se aproximam de muitas noções desenvolvidas no âmbito da fraseologia: têm pontos em comum com a noção de *motif*<sup>25</sup> criada por Mellet e Longrée, que a aplicaram ao estudo dos textos em latim (ver, por exemplo, Mellet e Longrée, 2012). O *motif* é definido como uma estrutura recorrente e linear, com uma função textual e discursiva. Nossa definição das rotinas é, ao mesmo tempo, mais estrita, na medida em que a função das rotinas é sobretudo *retórica* (a função discursiva caracterizará outros tipos de sequências, como já apresentado), e mais ampla, dado que a linearidade da expressão não nos parece uma característica chave. Exemplificando, as sequências *Contrairement à Parker (1990), nous...* [contrariamente a Parker (1990), nós...] e *Notre étude diffère de Parker (1990)* [nosso estudo difere do de Parker (1990)] recorrem a uma organização de actantes diferente (a fonte antes do agente no primeiro exemplo e de forma inversa no segundo), entretanto correspondem ao mesmo esquema para nós. Em contrapartida, parece essencial que os diferentes elementos presentes nessas estruturas estejam relacionados semântica e sintaticamente. Além disso, a noção de rotina retórica é bastante próxima daquela de *estruturas pré-fabricadas*, em particular da noção de *modelo de texto* desenvolvida por Gülich e Krafft (1997):

---

<sup>25</sup> N.T.: Do francês, *motif*: “[...] associação recorrente de X elementos do texto, seguindo uma determinada estrutura sintagmática, com função textual de encadeamento da informação.” (Sylvie Mellet, Dominique Longrée. *Légitimité d’une unité textométrique: le motif*. Anne Dister; Gérald Purnelle; Dominique Longrée. JADT 2012, 11èmes Journées internationales d’analyse statistique des données textuelles, Jun 2012, Liège, Belgique. Revue en ligne *Lexicometrica*, p. 717, 2012.).

Nomeamos *modelos de textos* estruturas pré-fabricadas de que nos servimos para a produção de certos tipos de texto [...]. Do ponto de vista da produção, o modelo é uma forma de receita que comporta indicações sobre as condições de emprego do texto: Onde? Em que momento? Em que condições?; Sobre sua estrutura: quais tipos de informações, eventualmente, em que ordem? Há prioridades a serem respeitadas?; e, com frequência, listas de expressões para utilização referencial (GÜLICH e KRAFFT, 1997, p. 258).

A noção de modelo de texto parece, contudo, mais global. As rotinas que descrevemos não se inserem necessariamente em uma estrutura linear: são enunciados estereotipados aos quais o locutor pode recorrer em qualquer parte do texto, mesmo que algumas funções sejam especialmente mais frequentes em certas partes, como a introdução. Por exemplo, as rotinas de prova científica (ex.: *comme on le voit dans le tableau 3* [como se vê no quadro 3]) são encontradas principalmente na parte metodológica, mas raramente nas introduções ou nas conclusões.

Se esse tipo de enunciado estereotipado foi, por fim, muito pouco estudado pela linguística, o TAL, sobretudo a área científica de extração de informação, propôs modelos mais sólidos. Esse domínio do TAL visa localizar, nos textos científicos, as partes que correspondem a conteúdos semânticos específicos. O estabelecimento de padrões de bases retórica e pragmática desenvolvidos a partir do TAL parecem mais estruturados (cf. Teufel, 1998; Sándor, 2007). O modelo de Teufel (1998), o AZ (*Argumentative Zoning*), inicialmente utilizado para o resumo automático, é ainda uma referência de TAL. Exemplificando, a expressão *to our knowledge* [até onde sabemos] é associada à função *gap introduction*; por sua vez, *when compared to* [quando comparado a] é associada à função de *comparação*. Porém, como observou Sándor (2007), as expressões levantadas por Teufel são apenas *pacotes lexicais* não organizados do ponto de vista sintático ou semântico: não são padrões fraseológicos, mas apenas listas de palavras, sem organização específica. Para preencher essa lacuna, Sándor propõe um modelo alternativo, o *concept matching*, que inclui uma representação sintática e semântica combinada com uma função discursiva, inspirada em parte por Swales (1990) (CARS): um conjunto de conceitos, realizados pelos marcadores lexicais e ligados pelas relações sintáticas, é associado a uma função retórica. Por exemplo, a função *conhecimento partilhado* integra os conceitos *passado, geral, pesquisador, atividade mental, objeto/resultado da atividade científica*, em que cada um se realiza de forma diferente no plano lexical. O Quadro 3, extraído de Sándor (2007), oferece uma versão esquematizada desse exemplo.

<i>background</i> [compartilhado]		<i>Knowledge</i> [conhecimento]		
<i>general</i> [geral]	<i>past</i> [passado]	<i>researcher</i> [pesquisador]	<i>mental activity</i> [atividade mental]	object result of research activity [objeto/resultado da atividade científica]
<i>most</i> [a maior parte]		<i>evolutionists</i> [evolucionistas]	<i>Consider</i> [considerar]	
	<i>recent</i> [recente]		<i>get a picture</i> [entender]	<i>findings</i> [descobertas]

Quadro 3: Exemplo da função discursiva conhecimento compartilhado (Sándor, 2007).

A abordagem proposta por Sándor parece de fato interessante, na medida em que sistematiza o léxico e as funções semânticas associadas à função retórica. Retomamos alguns de seus princípios ao sistematizar, por um lado, os papéis semânticos e, por outro, ao considerar os aspectos retóricos e enunciativos.

### 3.3. Frames semânticos

Propomos um modelo de representação, os *frames* semânticos, que permite sistematizar os enunciados estereotipados que correspondem a uma função retórica. Esses *frames* combinam as funções pragmáticas e retóricas com a dimensão semântico-enunciativa, assim como com aspectos lexicais e sintáticos (Cf. TUTIN, 2010b). A representação inclui três níveis:

- O nível retórico depende fortemente das estratégias discursivas em uso no gênero textual envolvido. No texto científico, é possível identificar, por exemplo, funções como *a demarcação em relação aos pares, a apresentação da problemática de pesquisa, a filiação científica*. É sem dúvida difícil propor aqui uma longa lista dessas funções, que se estabelece empiricamente através do estudo de *corpus*.<sup>26</sup>
- O nível semântico e enunciativo estabelece padrões com base nos esquemas abstratos associados às funções retóricas. Por exemplo, o *frame* da filiação científica articula um autor científico que toma de empréstimo um objeto científico desenvolvido por um autor fonte. Esse *frame* inclui os tipos semânticos de predicados, os papéis semânticos (no sentido de Fillmore) e os tipos semânticos dos participantes da situação enunciativa. Os tipos semânticos dos participantes são estreitamente ligados ao gênero e remetem aos protagonistas de um texto científico (o autor do texto, o leitor, a comunidade científica). Um *frame* conforme o de Fillmore (ex.: Fillmore *et al.*, 2003) permite estabelecer um padrão nesse nível. Entretanto, contrariamente aos *frames* semânticos do autor, nosso modelo não é cognitivo, mas sim semântico e textual: propõe uma representação abstrata de estereótipos textuais. Além disso, os tipos e papéis semânticos são estreitamente ligados ao gênero e só têm validade se associados a um modelo textual.

<sup>26</sup> Este trabalho está sendo realizado por nossa equipe.

- O nível sintático e lexical inclui os marcadores lexicais associados aos tipos semânticos, já as relações sintáticas, a seus papéis semânticos. Essa correspondência elucida a variedade lexical e sintática dos estereótipos textuais.

Tomemos o exemplo da filiação científica (Cf. GROSSMANN *et al.*, 2009), que pode ser considerada como uma função retórica clássica ligada ao texto científico, por meio da qual o autor indica sua pertença a uma linha, sua intenção de continuar um modelo teórico ou que faz uso de um modelo para o propósito de um estudo. Os exemplos seguintes, extraídos de textos científicos, dizem claramente respeito a essa noção de filiação:

- (1) *Cette étude reprend ici les travaux de Bunt (1995) sur...* [Este estudo retoma os trabalhos de Bunt (1995) sobre...]
- (2) *À la suite de (Lee 1986), nous ne considérerons pas que l'employeur...* [Dando seguimento a (Lee 1986), não consideraremos que o empregador...]
- (3) *Nous avons choisi d'utiliser le modèle des graphes conceptuels (Sowa, 1984)...* [Escolhemos utilizar o modelo de gráficos conceituais (Sowa, 1984)...]

Apesar da diversidade das realizações lexicais e sintáticas, um esquema comum da filiação pode ser associado a esses exemplos. Essa função retórica articula três participantes e um predicado genérico de *empréstimo*.

- um autor científico, em geral o autor do texto (*nous* [nós], *je* [eu]) ou, por metonímia, o objeto que ele constrói no texto (*cette étude* [este estudo], *ce travail* [este trabalho], *notre étude*<sup>27</sup> [nosso estudo]). Esse participante nem sempre se realiza na superfície por um marcador explícito. Por exemplo, em *le modèle reprend la notion de cadre développée par Fillmore* [o modelo retoma a noção de *frame* desenvolvida por Fillmore], não é possível, fora do contexto, determinar se é o modelo do autor ou de um terceiro.
- o objeto científico (*travaux* [estudos], *modèle* [modelo], mas também *terme* [termo], *idée* [ideia], *expression...* [expressão...]) é tomado emprestado ou utilizado pelo autor; trata-se das construções intelectuais que circulam na comunidade científica. O paradigma dos *objets scientifiques* é bastante amplo e depende evidentemente da disciplina. Em Grossmann et al. (2009), identificamos que os objetos científicos utilizados em Economia eram, com frequência, mais complexos (*analyse* [análise], *approche* [abordagem], *modèle* [modelo]) do que aqueles que apareciam em Linguística (*termes* [termos], *notion* [noção], *concept* [conceito]). Há casos em que a filiação não menciona especificamente o objeto utilizado, como no exemplo (2) acima: nesse caso, é a análise global do autor que é retomada.
- a fonte do empréstimo é uma referência, geralmente um autor ou autores associados a uma data (*Bunt (1995)*, [*Sowa, 1984*], grifos da autora), sendo

---

<sup>27</sup> Um dêitico de lugar pode também ser usado para remeter a um autor: *Le modèle développé ici reprend...* [O modelo desenvolvido aqui retoma...].

esta a convenção nos textos científicos. Em alguns casos, a fonte não se realiza pela convenção autor-data, mas por um adjetivo derivado (ex.: *nous reprenons la notion fillmorienne de cadre* [retomamos a noção fillmoriana de *frame*]).

- finalmente, um predicado verbal (*reprendre* [retomar], *utiliser* [utilizar]) ou nominal integrado a um sintagma preposicional (*à la suite de* [dando seguimento a], *dans la continuité de* [dando continuidade a]) indica o processo de empréstimo científico.

Os papéis semânticos (aqui chamados de maneira convencional *agente* [agente], *objet* [objeto] e *source* [fonte]) ligam o predicado “Empréstimo” aos objetos semânticos. A Figura 1 esquematiza essa análise.

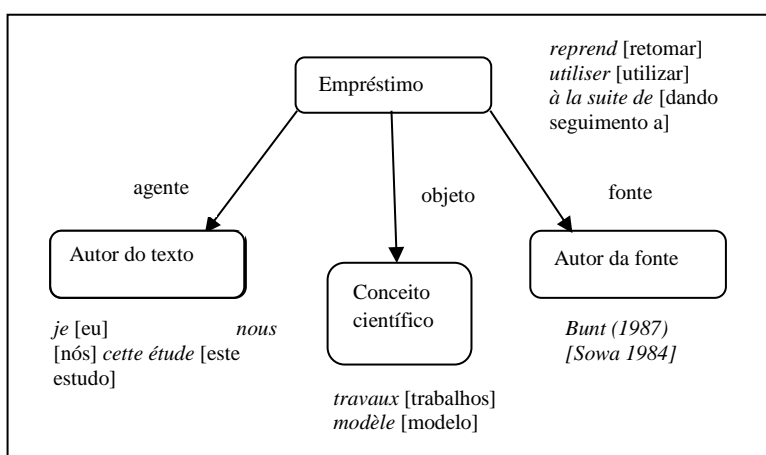


Figura 1: *Frame* semântico-retórico da filiação científica.

Do ponto de vista sintático-semântico, várias construções podem ser associadas a esse *frame*, como se observa na Figura 2. Essas construções são aqui representadas com a ajuda de árvores de dependência, do mesmo tipo que aquelas utilizadas na análise sintática para o projeto Scientext.

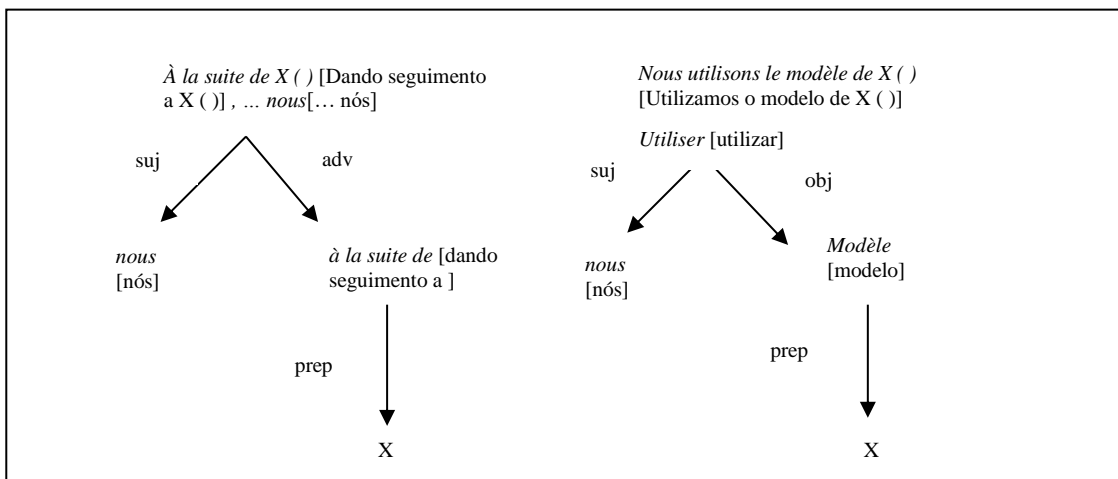


Figura 2: Exemplos de configurações sintáticas que correspondem ao *frame* da filiação científica.

Em frases como *à la suite de X...* [dando seguimento a X...], o conceito de *empréstimo* é introduzido pela preposição, enquanto em *nous utilisons le modèle de X* [utilizamos o modelo de X], o autor é o sujeito, e o conceito de *empréstimo* está ligado a um verbo. No projeto Scientext, cerca de 15 gramáticas locais foram implementadas de maneira simplificada seguindo esse princípio, as diferentes configurações sintáticas utilizam quase-sinônimos que remetem ao mesmo objeto semântico. Algumas relações sintáticas de dependência estão associadas a uma relação semântica: por exemplo, as construções ativas (ex.: *formuler une hypothèse* [levantar uma hipótese]), as construções passivas (ex.: *l'hypothèse est formulée* [a hipótese é levantada]) ou as passivas reduzidas (ex.: *hypothèse formulée* [hipótese levantada]) corresponderão a uma relação da ordem do “objeto profundo”, como proposto em Hagège e Roux (2003). Contudo, por razões de eficácia informática, versões simplificadas dessas gramáticas são hoje utilizadas no site Scientext<sup>28</sup>. As gramáticas baseiam-se na sintaxe do *corpus* analisado com ajuda do *software* Syntex, desenvolvido por Bourigault (2007). Essa linguagem de gramáticas locais, Conquest, foi desenvolvida por Kraif (2008) e formatada para a interface do Scientext e do Scienquest por Falaise (2013).

A Figura 3 apresenta uma versão simplificada da gramática da filiação. Observa-se que o autor pode ser um pronome de primeira pessoa ou *on* [sujeito indeterminado]. O mecanismo de empréstimo se realiza através de um verbo como *utiliser* [utilizar], *reprendre* [retomar], *choisir* [escolher], enquanto o objeto utilizado é um nome que pertence a uma lista de construtos científicos como *analyse* [análise], *méthode* [método] ou *approche* [abordagem]. A gramática elaborada é aplicada no *corpus* Scientext e permite extrair um conjunto de rotinas que exprimem a filiação.

```
// Frame da filiação
// Objeto
$Objet [$Objeto] = analyse [análise], approche [abordagem], concept [conceito], courant [corrente], critère
[critério], définition [definição], démarche [andamento], démonstration [demonstração], dénomination
[denominação], distinction [distinção], étude [estudo], facteur [fator], hypothèse [hipótese], mesure
[medida], méthode [método], modle [modelo], modélisation [modelização], notion [noção], prncipe
[princípio], procédure [procedimento], proposition [proposta], recherche [pesquisa], règle [regra], référence
[referência], remarque [observação], réflexion [reflexão], représentation [representação], statistique
[estatística], technique [técnica], terme [termo], terminologie [terminologia], théorème [teorema], théorie
[teoria], travail [trabalho], typologie [tipologia]
// Autor
$Auteur[$Autor] = nous[nós], je[eu], on[]
// Verbo de empréstimo
$Vemprunt [$Vempréstimo] =utiliser [utilizar], exploiter [explorar], employer [empregar], retenir [manter],
choisir [escolher], compléter [completar], poursuivre [proseguir], prolonger [prolongar], reprendre
[retomar], retenir [manter], étendre [ampliar], compléter [completar], suivre [seguir], Inscrire [inscrever],
recourir [recorrer]
Controle = <lemma=$Auteur,#1>&&<lemma=$Vemprunt,#2>&&<lemma=$Objet,#3> ::
(SUJ,#2,#1)(OBJ,#2,#3)
```

Figura 3: Uma gramática simplificada que representa o *frame* da filiação teórica.

<sup>28</sup> Objetivamos acrescentar diretamente ao *corpus* analisado essas relações semânticas de maneira a acelerar o processo de consultas no *site* que, neste momento, é muito demorado.



A gramática elaborada não é evidentemente perfeita – gera ruído e *silêncio*<sup>29</sup> –, mas permite extrair um conjunto de contextos pertinentes que ilustram esse tipo de rotinas, como os exemplos abaixo.

- (4) *Nous reprenons ici partiellement la typologie proposée dans Segond & Breidt (1995)...* [Retomamos parcialmente aqui a tipologia proposta em Segond & Breidt (1995)...] (Linguística – comunicação)
- (5) *Ce dernier potentiel ne permettant pas d'accéder à l'ensemble des énergies souhaitées, nous utiliserons donc la procédure de l'état de transition de Slater combiné au potentiel BP86...* [Como este último potencial não permite acessar o conjunto de energias desejadas, utilizaremos então o procedimento do estado de transição de Slater combinado ao potencial BP86...] (Eletrônica – tese)
- (6) *Nous utilisons le modèle des patrons (Blache et Rauzy, 2006) pour calculer la probabilité de chaque entrée du lexique global compte tenu du contexte syntaxique constitué par la séquence des formes orthographiques déjà saisies dans la phrase.* [Utilizamos o modelo de padrões (Blanche e Rauzy, 2006) para calcular a probabilidade de cada entrada do léxico global dado o contexto sintático constituído pela sequência de formas ortográficas já indicadas na frase.] (Processamento automático de linguagem – artigo)
- (7) *Nous retiendrons la règle de l'épenthèse (Klausenberger, 1974; Tranel, 1974), car elle est encore aujourd'hui envisagée, cependant sous une forme différente (Côté, 2005)* [Nós manteremos a regra da epêntese (Klausenberger, 1974; Tranel, 1974), pois ela é ainda hoje considerada, mas de uma forma diferente (Côté, 2005)] (Linguística – tese)

Futuramente desejamos melhorar a implementação desses *frames*, que é ainda muito rudimentar para dar conta da diversidade dos padrões, desenvolvendo a análise semântica acima. Desejamos, da mesma forma, desenvolver o estudo das funções retóricas de maneira a aumentar a apreensão dos fenômenos. Esse trabalho não pode, entretanto, ser automatizado, contrariamente aos outros tipos de sequências lexicalizadas que podem ser extraídas semiautomaticamente, e requer um profundo conhecimento do texto científico. Por fim, em colaboração com nossos colegas da didática e especialistas no processamento automático dos *corpora*, desejamos refletir sobre as aplicações didáticas simples que utilizem os resultados dos recursos que poderão ser integrados às atividades de observação e aos exercícios.

## 5. Conclusão

A fraseologia transdisciplinar dos textos científicos é diversa. A abordagem que propusemos aqui é essencialmente funcional e distingue quatro grandes tipos de sequências polilexicais segundo sua função predominante: (a) colocações e expressões fixas

---

<sup>29</sup> O ruído corresponde às extrações não apropriadas, já o silêncio concerne aos exemplos presentes no texto que não são detectados pelo recurso.

de função referencial que remetem primordialmente aos procedimentos e aos objetos científicos; (b) marcadores discursivos que estruturam o discurso ou a argumentação; (c) sequências polilexicais de função interpessoal que remetem aos traços da situação enunciativa e à presença do locutor/autor; (d) rotinas retóricas que remetem a funções retóricas estereotipadas do gênero em questão. É provável que a tipologia esboçada aqui não esgote todas as possibilidades, além disso, sua aplicação em um âmbito descritivo, o que estamos em vias de realizar, certamente não se dará sem dificuldades<sup>30</sup>.

Voltamo-nos, em seguida, para as rotinas retóricas que, pela diversidade de suas realizações lexicais, situam-se no limite da fraseologia, mas apontam para a retórica estereotipada dos textos científicos. O estabelecimento de padrões proposto, inspirado livremente nos *frames* semânticos de Fillmore, foi implementado de forma simplificada com a ajuda de gramáticas locais e precisa ainda ser validado com um número maior de dados. Por último, ainda que o modelo tenha sido aplicado, até o momento, principalmente aos textos científicos – em que as diferentes funções retóricas se realizam em uma língua fortemente rotinizada –, ele pode ser, ao que tudo indica, utilizado para a descrição de outros socioletos, como as línguas profissionais ou o discurso político.

Na continuação do projeto Scientext, desejamos desenvolver os recursos fraseológicos conforme a tipologia proposta para o estudo epistemológico dos textos científicos e para aplicações didáticas.

## Referências

- BIBER D., *University Language: a corpus-based study of spoken and written registers*, Amsterdam/Philadelphia, John Benjamins, 2006.
- BOURIGAULT D., *Un analyseur syntaxique opérationnel : SYNTEX*, mémoire de HDR, Toulouse, 2007.
- BURGER H., *Phraseologie: eine Einführung am Beispiel des Deutschen*, E. Schmidt, 2007.
- CAVALLA C., « Les écrits universitaires des étudiants étrangers : quelles normes présenter ? », BERTRAND O., SCHAFFNER I. (dir.), *Quel français enseigner ? La question de la norme dans l'apprentissage/enseignement*, Paris, Editions Ecole Polytechnique, 2010, p. 203-214.
- CORMAN J., *Extraction d'expressions polylexicales sur corpus arboré*, Mémoire de master 2 recherche, Sciences du langage, Grenoble, Université Grenoble 3 – Stendhal, 2012.
- COXHEAD A., “The Academic Word List: A Corpus-based Word List for Academic Purposes”, KETTERMAN B., MARKS G. (ed), *Teaching and Language Corpora (TALC) 2000 Conference Proceedings*, Atlanta, Rodopi, 2002, p. 73-89.

---

<sup>30</sup> Em colaboração com Thi Thu Hoai Tran, doutorando, e no âmbito do projeto ANR-Contint Termith com Marie-Paule Jacques, Cristelle Cavalla e Sylvain Hatier.

- DIWERSY S., KRAIF O. « Le Lexicoscope : un outil pour l'analyse de profils combinatoires et l'extraction de constructions lexico-syntaxiques », *Actes de la conférence conjointe JEP-TALN-RECITAL 2012*, volume 2, juin 2012, Grenoble, p. 399-406.
- FALAISE, A. « Exploitation linguistique de corpus arborés d'écrits scientifiques à l'aide du logiciel ScienQuest », TUTIN, A. ; GROSSMANN, F. (dir.), *L'écrit scientifique : du lexique au discours*, Rennes : PUR, 2013, p. 123-141.
- FILLMORE C. J., JOHNSON C. R., PETRUCK M. R. "Background to Framenet", *International Journal of Lexicography*, n°16(3), 2003, p.235-250.
- FLØTTUM K., DAHL T., KINN T. *Academic Voices. Across languages and disciplines*, Amsterdam/Philadelphia, John Benjamins, 2006.
- FÓNAGY I., « Figement et changement sémantiques », M. MARTINS-BALTAR (ed), *La Locution entre langue et usages*, Fontenay-Saint-Cloud : ENS Editions, 1997, p. 131-164.
- GLEDHILL C., *Collocations in science writing*, Language in performance, 22, Tübingen, Gunter, 2000.
- GRANGER S., PAQUOT M., "Disentangling the phraseological web", Granger S., Meunier F, *Phraseology: An interdisciplinary perspective*, Amsterdam/Philadelphia, John Benjamins, 2008.
- GROSSMANN F., TUTIN A. (dir.), *Les collocations : analyse et traitement*, Travaux et recherches en linguistique appliquée, Amsterdam, de Werelt, 2003.
- GROSSMANN F., TUTIN A. GARCIA DA SILVA P., « Filiation et transferts d'objets scientifiques dans les écrits de recherche », *Pratiques*, n°143-144, 2009, p. 187-202.
- GÜLICH E., KRAFFT U., « Le rôle du préfabriqué dans les processus de la production discursive », MARTINS-BALTAR M. (dir.), *La locution entre langue et usages*, Paris, ENS éditions. Fontenay-St Cloud, 1997, p. 241-276.
- HAGEGE C., ROUX C. « Entre syntaxe et sémantique : Normalisation de la sortie de l'analyse syntaxique en vue de l'amélioration de l'extraction d'information à partir de textes », *Actes de TALN, TALN 2003*, Batz-sur-Mer, 11-14 juin 2003.
- HALLIDAY M.A.K., *Introduction to functional grammar*, 2nd ed., London, Edward Arnold, 1994.
- HOWARTH P.A., *Phraseology in English Academic Writing*. Tübingen, Max Niemeyer, 1996.
- HYLAND K., *Disciplinary discourses: social interactions in academic writing*, London, Longman, 2000.
- HYLAND K., *Metadiscourse. Exploring Interaction in Writing*, London, Continuum, 2005

KOCOUREK R., *La langue française de la technique et de la science : vers une linguistique de la langue savante*, Wiesbaden, O. Brandstetter, 1991.

KRAIF O., « Comment allier la puissance du TAL et la simplicité d'utilisation ? l'exemple du concordancier bilingue ConcQuest », *Actes des 9ème Journées d'analyse statistique des données textuelles, JADT 2008*, Lyon, Presses universitaires de Lyon, 2008, p. 625-634.

LAFON P., SALEM A. « L'Inventaire des segments répétés d'un texte », *Mots*, n°6, 1983, p.161-177.

MANGIANTE J.-M., PARPETTE C., *Le français sur objectif universitaire*, Grenoble, Presses Universitaires de Grenoble, 2011.

MEL'CUK I. « Phrasèmes dans le dictionnaire », MEJRI S., ANSCOMBRE J.CL (dir.), *Le figement linguistique : la parole entravée*, Honoré Champion, Paris, 2011, p. 41-61.

MELLET S., LONGREE D., « Légitimité d'une unité textométrique : le motif », *Lexicometrica, Actes de JADT 2012*, 2012, p. 712-728.

PECMAN M., *Phraséologie contrastive anglais-français : analyse et traitement en vue de l'aide à la rédaction scientifique*, Thèse de doctorat, Université de Nice Sophia Antipolis, 2004.

PECMAN M., « Approche onomasiologique de la langue scientifique générale », *Revue Française de Linguistique Appliquée*, Vol XII, 2007-2, p. 79-96.

PERY-WOODLEY M.-P., *Une pragmatique à fleur de texte : approche en corpus de l'organisation textuelle*, Mémoire d'habilitation à diriger des recherches, Carnet de grammaire, Rapport n° 8, 2000.

SÁNDOR Á., "Modeling metadiscourse conveying the author's rhetorical strategy in biomedical research abstracts", *Revue Française de Linguistique Appliquée*, XII-2, 2007, p. 97-109.

SERETAN V., *Syntax-Based Collocation Extraction*, Dordrecht, Springer, 2011.

SIEPMANN D., « Les marqueurs de discours polylexicaux en français scientifique », *Revue française de linguistique appliquée*, 2/2007, (Vol. XII), p. 123-136.

SWALES J. M., *Genre Analysis: English in academic and research settings*, Cambridge, Cambridge University Press, 1990.

TEUFEL S. « Meta-discourse markers and problem-structuring in scientific articles », *ACL 1998 Workshop, Discourse Structure and Discourse Markers*. Montreal, Somerset/New Jersey, 1998, p. 43-49.

TUTIN A. [2007a] (dir.), *Revue Française de Linguistique Appliquée, volume XII-2, Lexique et écrits scientifiques*, 2010.

TUTIN A. [2007b], « Collocations du lexique transdisciplinaire des écrits scientifiques :

annotation et extraction des propriétés linguistiques dans la perspective d'une application didactique », *Cahiers de l'Institut de linguistique de Louvain*, n°Vol. 31, n° 2-4, 2007, p. 247-262.

TUTIN A. [2010b], "Showing phraseology in context: an onomasiological access to lexicogrammatical patterns in corpora of French scientific writings", Granger S., Paquot M. (ed), *Proceedings of eLex 2009, Louvain-la-Neuve, 22-24 October 2009*, Cahiers du Cental, Louvain, Presses universitaires de Louvain, 2010, p. 302-312.

Como citar este texto (ABNT):

TUTIN, A. Tradução de Anna Carolina Teles. A fraseologia transdisciplinar dos textos científicos: das colocações às rotinas retóricas. *Cadernos de Tradução*, Porto Alegre, n. 43, jul/dez, p. 170-190, 2018.