



## CogniTExtes

Revue de l'Association française de linguistique cognitive

Volume 19 | 2019

Corpora and Representativeness

---

# Énonciation, corpus et représentativité : le cas de « along »

Graham Ranger

---



### Édition électronique

URL : <http://journals.openedition.org/cognitextes/1517>

DOI : 10.4000/cognitextes.1517

ISSN : 1958-5322

### Éditeur

Association française de linguistique cognitive

### Référence électronique

Graham Ranger, « Énonciation, corpus et représentativité : le cas de « along » », *CogniTExtes* [En ligne], Volume 19 | 2019, mis en ligne le 31 juillet 2019, consulté le 01 juillet 2019. URL : <http://journals.openedition.org/cognitextes/1517> ; DOI : 10.4000/cognitextes.1517

---

Ce document a été généré automatiquement le 1 juillet 2019.



*CogniTExtes* est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.

---

# Énonciation, corpus et représentativité : le cas de « along »

Graham Ranger

---

## 1. Introduction : la représentativité du corpus vue du côté de la linguistique énonciative

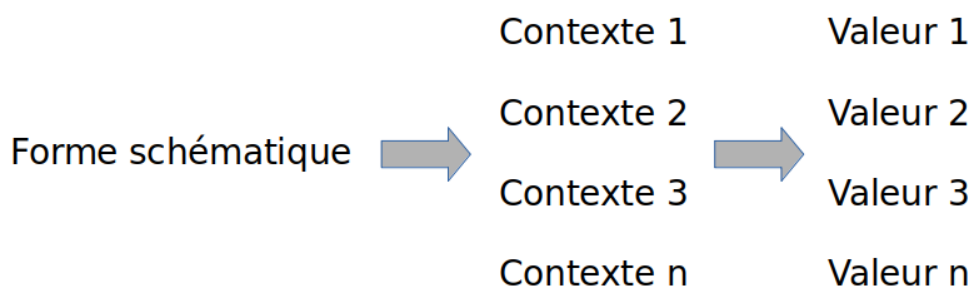
- 1 Le chercheur qui souhaite travailler sur le langage est confronté au choix difficile entre deux chemins également hasardeux, dès lors qu'il souhaite définir précisément son objet d'étude. Soit, il s'appuie sur de l'introspection, puisant dans ses propres compétences langagières les données sur lesquelles il va ensuite théoriser, soit, il se base sur des données attestées, rassemblées le plus souvent sous forme de corpus. Les écueils des deux démarches sont bien connus. Le recours aux « données introspectives » – la formulation relève déjà de l'oxymore – suppose chez le linguiste une forme de compétence universelle qui lui permettrait, à partir d'énoncés et de situations imaginés, avec plus ou moins de vraisemblance, par un seul sujet, de théoriser sur des contraintes d'ordre générique, s'appliquant à tout sujet, et de faire abstraction des fâcheuses « erreurs de performance »... Il est difficile de voir comment une telle méthode pourrait représenter autre chose que la compétence langagière d'un sujet unique, s'exerçant sans ancrage situationnel authentique.
- 2 Le recours à des corpus extérieurs à la psychologie du linguiste-chercheur présente déjà l'avantage de l'objectivation des données, mais comporte d'autres inconvénients. Considérons d'abord la constitution d'un corpus généraliste, qui vise à fournir d'une langue donnée une image aussi fidèle, et aussi variée que possible. En admettant qu'il n'y ait pas un seul registre qui soit représentatif de l'ensemble d'une langue, on procédera par échantillonnage, retenant une quantité déterminée de mots dans un éventail de registres préalablement défini. La question qui se pose alors est celle de l'importance relative de chaque échantillon. Admettons qu'on retient un million de mots, par exemple, pour chaque registre : conversation familière, entretien, récits littéraires, contrats juridiques, science-fiction, poésie... On pourra objecter que la conversation familière, qui

occupe une partie très significative de notre activité langagière au quotidien, se trouve sous-représentée par rapport à la poésie, ou par rapport aux contrats juridiques. Une telle objection se place du côté de la réception, en prônant l'utilisation d'un corpus qui soit représentatif de la langue telle qu'on y est exposé, ou plutôt telle qu'y est exposé un sujet locuteur lambda. Le problème de la représentativité se trouve ainsi déplacé de la langue-type vers le locuteur-type. Les choses ne sont pas plus satisfaisantes si l'on se place du côté de la production, cependant. En effet, si on veut bien admettre que le registre le plus présent quantitativement est celui de la conversation familière, faut-il pour autant baser un modèle du langage sur ce seul mode énonciatif ?

- 3 Le recours à des corpus spécialisés, limités à la représentation d'un registre linguistique, permet de sortir des problèmes qu'implique la tentative de représenter toute une langue, mais se heurte à d'autres obstacles. En effet, travailler sur un corpus spécialisé (le langage des contrats juridiques, des articles de médecine, des tweets...) restreint nécessairement les conclusions que l'on pourra tirer de sa recherche au seul registre étudié. On peut essayer de caractériser les spécificités d'un corpus spécialisé vis-à-vis d'un corpus généraliste, mais on retrouve alors les problèmes déjà mentionnés dans la définition de celui-ci.
- 4 Ce sont là des problèmes bien connus auquel le présent travail ne prétend pas apporter de réponses<sup>1</sup>. Plutôt, nous proposons de considérer la représentativité du corpus du point de vue des objectifs que se fixe le linguiste, en fonction de son cadre théorique.
- 5 On peut considérer que la Théorie des Opérations Prédicatives et Énonciatives (TOPÉ) n'a pas mené de véritable réflexion de fond sur le statut du corpus. Culioli insiste à de nombreuses reprises sur la nécessité de travailler sur des énoncés authentiques en contexte, dont la situation d'énonciation est explicitée<sup>2</sup>. Et cependant les études s'inscrivant dans le cadre de la TOPÉ continuent de s'inspirer pour l'essentiel de corpus littéraires, exploités comme vivier d'exemples authentiques, mais sans réflexion épistémologique sur les écarts possibles entre les spécificités textuelles du récit littéraire et le phénomène linguistique qu'on voudrait décrire, au-delà des frontières d'un seul genre. C'est une approche qui s'explique par des raisons techniques (le corpus littéraire est souvent disponible sous forme numérisée) et sociologiques (le corpus littéraire est une variété de langue valorisée, à la fois par la société, et par les linguistes, souvent issus eux-mêmes de formations littéraires). Il n'empêche qu'une telle approche privilégie, la plupart du temps sans mise en perspective, un mode énonciatif historiquement et culturellement très spécifique (le récit de fiction réaliste)<sup>3</sup>. Par ailleurs, la Théorie des Observables, préconisée par la TOPÉ, implique de travailler sur des énoncés authentiques, mais aussi de les manipuler, de manière contrôlée, afin d'en tester les limites d'acceptabilité, de mise en contexte, etc.<sup>4</sup> Là encore, on pourra interroger la pertinence, ou du moins la compatibilité de la démarche avec une approche basée sur corpus, car travailler sur un énoncé manipulé revient, nécessairement, à travailler sur une fabrication de linguiste. La difficulté, c'est que le corpus ne peut nous renseigner que sur ce qui existe, tandis que la description d'un phénomène linguistique, d'un marqueur, etc. nécessite la recherche des limites aux emplois. Les impossibilités sont, dans cette optique, aussi révélatrices que les possibilités.
- 6 Culioli définit les objectifs de la linguistique de la manière suivante : « Je dirai que la linguistique a pour objet l'activité du langage appréhendée à travers la diversité des langues naturelles (et à travers la diversité des textes, oraux ou écrits) » (1991 : 14). La TOPÉ, élaborée par Culioli et des chercheurs associés, décrit la langue à l'aide d'un

appareil métalinguistique robuste, composé d'un nombre réduit d'opérations et d'opérandes. Les marqueurs d'une langue sont caractérisés en termes d'une forme schématique (FS). Les différentes valeurs situées des marqueurs sont le résultat d'un calcul de paramétrage de cette forme schématique, en fonction du contexte, selon un processus qui est représenté grossièrement en Figure 1.

Figure 1 : Paramétrage de la forme schématique en contexte.



- 7 Nous estimons que le travail sur « la diversité des textes, oraux ou écrits » et sur les différentes valeurs contextualisées des marqueurs linguistiques, oblige à faire appel à des corpus multi-genres. Par ailleurs, dès lors que le corpus en question est suffisamment étendu, suffisamment varié et suffisamment bien annoté, il devient possible de quantifier, par le biais des mesures d'associativité, les paramétrages contextuels des marqueurs, ce qui informe, et oriente l'élaboration d'une forme schématique, obviant en partie aux inconvénients que peuvent présenter les manipulations.
- 8 La suite de cet article est consacrée à une étude énonciative du marqueur *along* basée sur des données issues du British National Corpus, interrogé via l'interface BNCweb<sup>5</sup>.

## 2. Etude de cas : *along* et *alongside*

- 9 La linguistique cognitive a souvent accordé une place privilégiée à l'étude des prépositions, aux rapports entre repérages spatiaux et autres types de relations, etc.<sup>6</sup> Le marqueur *along* est caractérisable comme *préposition*, et comme *particule*, mais n'a pas, à notre connaissance, fait l'objet de travaux conséquents, quel que soit le cadre théorique. Notre choix de travailler sur le marqueur *along* s'est présenté à l'occasion de difficultés constatées lors d'ateliers de traduction, dans la traduction de la forme française *le long de*. Le texte source à traduire est donné dans (1).

1) Une jeune femme apparaissait, dans le cadre étroit de sa vision. Deux, plutôt : deux femmes. Aussitôt après la première, qui marchait sur la chaussée, le long des voitures en stationnement vers Zapata, une deuxième jeune femme se montrait, en effet, s'avancant sur le trottoir, elle.  
[Source : Jorge Semprun, *Netchaïev est de retour* (1987).]

- 10 Ici il s'agit donc du passage « qui marchait sur la chaussée, le long des voitures en stationnement », qui a donné lieu à au moins trois traductions (1a), (1b), (1c)<sup>7</sup>.

- 1a) → ? ? who was walking ... along the parked cars  
 1b) → who was walking ... alongside the parked cars  
 1c) → who was walking ... along the row of parked cars

- 11 Si la traduction de *le long de* par *alongside*, ou encore l'étoffement par *along the row of*, aboutissent à des textes cibles idiomatiques, il n'en va pas de même pour la traduction par calque, *along the parked cars*, qui se caractérise par une forme d'incompatibilité entre la préposition *along* et son complément *the parked cars*. Afin de vérifier l'intuition d'inacceptabilité que matérialisent les points d'interrogation, et de cerner la différence entre *along* et le composé *alongside*, d'abord, nous choisissons d'interroger le BNC, à la recherche de lemmes verbaux à gauche et de lemmes nominaux à droite des deux prépositions. Les résultats sont présentés dans les tableaux (1) et (2)<sup>8</sup>.

Tableau 1 : Résultats d'une requête BNCweb pour les lemmes verbaux de *along* et *alongside* dans une fenêtre de trois mots à gauche.

Rank	along	Log-Likelihood	alongside	Log-Likelihood
1	come	6405.356	work	1044.2271
2	walk	5436.1998	run	408.7615
3	go	5037.4402	stand	239.4661
4	run	2122.8921	fight	225.5931
5	drive	992.7869	play	224.2257
6	move	914.4928	sit	217.3745
7	travel	641.4612	come	149.1628
8	hurry	597.6246	rank	142.7978
9	bring	544.8218	set	126.8625
10	tag	464.6472	walk	125.9929

Tableau 2 : Résultats d'une requête BNCweb pour les lemmes nominaux de *along* et *alongside* dans une fenêtre de trois mots à droite

Rank	along	Log-Likelihood	alongside	Log-Likelihood
1	line	9408.7139	other	170.2628
2	road	5505.7895	road	146.3086
3	coast	3055.2873	river	144.0279
4	path	2766.8444	jetty	80.2246
5	corridor	2403.8475	stream	59.8731
6	edge	1389.7126	ermine	57.6525
7	street	1267.3222	wall	53.4595
8	length	1207.23	loch	50.9383
9	side	1205.8376	fence	44.9818
10	river	1198.1442	prost	44.9391

- 12 Ces premiers résultats appellent plusieurs commentaires. D'abord, on relève une différence de *nature* entre les collocations nominales et verbales de *along* et *alongside* : pour ce qui est des verbes, seuls *come* et *walk* figurent parmi les dix premières collocations des deux marqueurs ; pour ce qui est des noms, seuls *road* et *river* figurent parmi les dix

premières collocations. Ensuite, on relève une différence de *degré* dans l'indice de probabilité logarithmique entre les deux marqueurs<sup>9</sup>. Enfin, on note une différence de degré également entre les indices de probabilité logarithmique pour les contextes droite ou gauche des deux marqueurs. On peut déjà formuler quelques hypothèses :

- 13 *Along* semble imposer de fortes contraintes de sélection sur les termes susceptibles de le précéder ou de le suivre, favorisant des verbes de déplacement, à gauche, et des noms possédant une propriété saillante de /LONGUEUR/ à droite<sup>10</sup>.
- 14 *Alongside* semble autoriser aussi bien des verbes d'état que des verbes de déplacement, et ne pas imposer de contraintes de sélection très caractérisées sur son contexte droit.
- 15 Ces hypothèses s'alignent sur nos premières intuitions concernant l'acceptabilité respective des traductions (1)-(3). Concrètement, l'impossibilité de ? ? *along the parked cars* vient de ce que la /LONGUEUR/ n'est pas *a priori* une propriété saillante des voitures en stationnement. La traduction *alongside the parked cars* est en revanche possible, car *alongside* n'impose pas des contraintes de sélection *a priori* mais *construit* pour sa cible la propriété /LONGUEUR/. Quant à *the row of parked cars*, le terme *row* permet à la fois de délimiter une certaine quantité de voitures en stationnement, mais de manière plus critique, il fournit une cible en longueur pour la préposition *along*.

### 3. La modélisation des prépositions et des particules dans la TOPÉ

- 16 Jusqu'à présent, nous avons simplement utilisé les requêtes sur corpus pour vérifier, par le biais de probabilités d'associativité, notre intuition concernant les possibilités de traduction de *le long de* dans le texte de Semprun. Nous allons, dans cette section, mettre en rapport ce type de constat essentiellement quantitatif, avec la modélisation de la forme schématique telle que la TOPÉ la pratique.
- 17 Posons tout d'abord qu'une préposition est la trace d'une opération de localisation entre deux termes, tel que le premier terme gagne en détermination du fait de sa localisation relativement au second. En d'autres mots, l'on a affaire à une opération de repérage entre un repéré à gauche et un repère à droite. En symbolisant le repéré par *x*, le repère par *y* et l'opération de repérage par le métaopérateur  $\in$ , on admettra que l'énoncé *Jane walked along the road*, correspond à une suite repéréopérationrepère (2)<sup>11</sup>.

2) Jane [came out of the tube station and] walked along the road. ABW 481 <sup>12</sup>
---

2a) Jane walked ( <i>x</i> ) along ( <i>préposition</i> ) the road ( <i>y</i> )
---

2b) $\langle x \in y \rangle$ or « <i>x</i> is located relative to <i>y</i> »
---

- 18 Nous avons ainsi défini une relation selon laquelle *x* est déterminé par sa relation avec *y*. En même temps, le choix de *y* n'est pas libre ; il est contraint par la relation qui relie *y* à *x*. Ainsi, dans *Jane walked along the road*, le repère *the road* est membre d'une classe de repères possibles, contrainte par la séquence à sa gauche *Jane walked along...* On peut représenter cette relation de présélection par la formule en (3), où les parenthèses vides représentent

une classe de termes repères potentiels pour  $x$  et où les indices rendent compte de l'imbrication complexe des relations<sup>13</sup>.

$$3) <_1 y \in <_2 ( ) y >_1 \ni x >_2$$

- 19 En combinant les deux formules, (2b) et (3), on aboutit à une formule complexe (4) qui représente à la fois la localisation de  $x$  relativement à  $y$  et la présélection de  $y$  en fonction de la nature de sa relation avec  $x$ <sup>14</sup>.

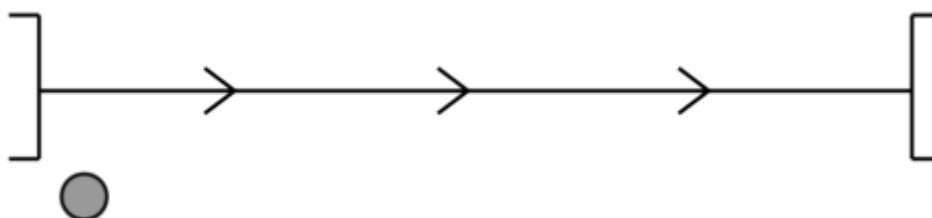
$$4^{\circ} <_1 x \in <_2 y >_1 \in <_3 ( ) y >_2 \ni x >_3$$

- 20 Ceci représente le cas des prépositions, mais lorsque le repère  $y$  est implicite, alors *along* est dit « particule », ce qui veut dire qu'il représente à la fois la relation de repérage et le repère<sup>15</sup>. Ainsi, dans (5), *when he's walking along* le repère, non spécifié, est déterminé de façon circulaire : *when he's walking along where he's walking*.

5) He trips up quite a bit anyway when he's walking along. ASH 849

- 21 La forme schématique d'un marqueur est étroitement liée aux latitudes de variation qui lui sont associées. Comme l'écrit Paillard : « L'identité du mot est [...] indissociable de la prise en compte de la séquence [...] où il apparaît, et il importe dès lors, pour un mot donné, de définir les différents contextes où ce mot s'inscrit. Cela revient à dire que le mot se définit par ses usages (nés de l'interaction entre le mot et ses contextes) : l'identité du mot est dans sa variation. » (1998 : 16). S'agissant de *along* les latitudes de variation sont, comme nous l'avons observé, des processus de déplacement ( $x$ ) et des repères évoquant la propriété /LONGUEUR/.
- 22 Nous proposons alors la forme schématique suivante, pour *along* :
- 23 Le marqueur *along* localise un terme  $x$  dynamiquement relativement à un espace non borné séquentiel  $y$  de telle façon qu'à  $T_0$ ,  $x$  est à  $Y_0$ , à  $T_1$ ,  $x$  est à  $Y_1$ , etc.
- 24 La Figure 2 représente une tentative de visualiser les spécificités de *along* ou l'espace séquentiel non borné  $y$  est représenté par un trait fléché, tandis que le repérage dynamique de  $x$  par rapport à  $y$  est représenté par un trait avec une flèche finale. Le terme  $x$  est mobile, le terme  $y$  est statique<sup>16</sup>.

Figure 2 : Représentation graphique de la forme schématique postulée pour *along*



- 25 Il s'agit là d'un dispositif topologique abstrait qui doit être compatible avec la diversité des valeurs situées en contexte. Nous n'entrerons pas ici dans la représentation de *alongside*. Notons cependant qu'à la différence de *along*, 1) *alongside* n'est pas nécessairement dynamique et ne requiert par conséquent pas de processus de déplacement ; 2) *alongside* construit chez le repère *y* une propriété de longitudinalité tandis que celle-ci est un *prérequis notionnel* pour *along*.
- 26 Le paramètre le plus critique, dans la modélisation des différentes valeurs en contexte de *along* est la configuration du terme repère *y* et, dans une moindre mesure, celle du terme repéré *x*. Ce sont ces deux latitudes de variation qui permettent à *along* de construire des valeurs spatiales, temporelles ou argumentatives / subjectives, comme nous allons le montrer, dans les sections suivantes, à nouveau à la lumière de requêtes ciblées sur le BNC.

## 4. Configurations spatiales de *along*

- 27 Les configurations spatiales sont, sans surprise, les mieux représentées dans le BNC. Parmi un échantillon aléatoire de 1000 occurrences de *along*, 63,5 % ont été catégorisées comme « spatial ».<sup>17</sup> Dans ce cas de figure, le terme *x* renvoie à un déplacement spatial, et, surtout, *y* renvoie, ou bien à une entité concrète dotée de /LONGUEUR/ dans le cas où *y* est instancié, ou bien à une trajectoire reconstruite, dans le cas où *y* n'est pas instancié et où *along* revêt un fonctionnement dit de « particule ».
- 28 (5) et (6) représentent le cas de *along* préposition, où *y* est représenté respectivement par *the coast* et *the train of fish vans*.

5) At first, sailors used to sail along the coast [...] G25 899

6) He slowly made his way along the train of fish vans. CE9 705

- 29 Dans (7) *along* est « particule », et *y* doit être reconstruit, ici à partir des indices dans le contexte gauche *accompany him... fell in beside him...* :

7) Hoomey was not anxious for Nails to accompany him, but Nails fell in beside him, slouching along with his hands in his pockets, his face sullen. AT4 1808

- 30 Dans de nombreux cas, la particule *along* est complétée d'une expansion prépositionnelle qui lui sert de repère à son tour comme dans (8), où la trajectoire à laquelle renvoie *along* est déterminée par *to Ballechin*, etc.

8) They tramped along to Ballechin among the ferns and sheltering oaks, smelling the cheesy smell of cows among the mist. A0N 903



## 5. Configurations temporelles de *along*

- 31 4,6 % de notre échantillon de *along* relève de configurations temporelles, où l'espace séquentiel défini par ce marqueur évoque un mouvement temporel, une série d'étapes, etc. Ainsi en (9) et (10) les déterminations contextuelles de *three years ago* de *step* ou de *it takes a long time... consequences* respectivement permettent de considérer qu'on a affaire à un passage à travers le temps.

7) Forest envisage the return of Webb, who left them in a £ 1.5 million deal three years ago, as a crucial step along the road to safety. CEP 5006

8) The problem with weight loss is that it takes a long time to get what you want. It is necessary, therefore, to arrange additional, pleasing consequences along the way. AD0 1282-1283

- 32 En opposition à la thèse d'une configuration temporelle de *along* opérée par le contexte, on pourrait objecter que, dans le cas de (7) et (8), c'est le terme *y*, *road* et *way*, respectivement qui est employé de manière non littérale. C'est, de fait, une démarche typique en linguistique cognitive que de privilégier une catégorie d'emplois concrets qui s'étend par métaphore vers d'autres, plus abstraits. Un inconvénient d'une telle approche réside en la difficulté à contrôler le recours à la métaphore. Dans la TOPÉ une valeur peut être plus fréquente, plus saillante, sans pour autant qu'elle soit intrinsèquement prioritaire sur d'autres. La description d'un marqueur est ainsi sous-déterminée, les valeurs en contexte résultant à chaque fois d'un paramétrage calculable<sup>18</sup>.
- 33 Un argument en faveur de cette approche est fourni par les valeurs temporelles de *along* en tant que particule en (9) et (10).

9) Theories and explanations must emerge from the work as it goes along. CMF 892

10) This is an excellent interactive tutorial as you can try each of the described commands as you go along. HAC 7833

- 34 Dans ces occurrences, *along* figure dans une circonstancielle de temps. Le terme repère *y* n'est pas explicite et se reconstruit comme renvoyant à l'espace temporel non borné constitué par le contexte gauche, *the work* et *tutorial* respectivement. En l'absence d'un terme *y* métaphorisable on doit considérer qu'il s'agit bel et bien d'un paramétrage temporel de *along*<sup>19</sup>.

## 6. Configurations argumentatives et subjectives de *along*

- 35 Comparons les exemples (11) et (12).

11) Join up and fold along lines. A70 1871

12) By now my mind was working along similar lines. A89 605

- 36 Dans le (11), les lignes évoquées sont des entités concrètes, matérialisées, et l'emploi fait de *along* relève par conséquent d'une configuration spatiale. Dans le (12), en revanche, le terme *lines* renvoie à un cheminement de pensée. Il s'agit d'une illustration de la possibilité de configuration argumentative du marqueur, où l'espace séquentiel défini par *y* correspond à un mouvement argumentatif ordonné, tel que cause et conséquence, prémisse et conclusion, etc. qui fonctionne comme repère pour *x*.
- 37 Si on veut bien reconnaître l'existence de ce cas de figure, on peut supposer qu'il est associé à certaines configurations textuelles, potentiellement observables dans le corpus. On se souviendra que le lemme *line* figurait en première position parmi les lemmes nominaux dans le contexte droit de *along* (Tableau 2) avec une probabilité logarithmique particulièrement élevée. La construction *along... lines* n'est pas nécessairement argumentative, comme le montre (11), mais elle paraît fortement orientée vers ce type d'interprétation. Pour vérifier cette hypothèse, nous pouvons rechercher les lemmes verbaux dans une fenêtre de trois mots à gauche du terme *along* suivi dans les trois mots à droite du lemme *line*. Les résultats de cette requête sont présentés dans le tableau 3.

Tableau 3 : Résultats d'une requête pour les lemmes verbaux à trois mots à gauche de la suite *along* suivi du lemme *line* dans une fenêtre de trois mots

Rank	Lemma	Log-Likelihood
1	run	203.39
2	develop	177.3068
3	think	167.9685
4	divide	132.5282
5	organise	82.7442
6	be	57.6017
7	work	55.6288
8	proceed	55.5744
9	organize	51.1413
10	go	44.8757

- 38 Les probabilités logarithmiques sont relativement faibles, si on les considère lemme par lemme. On remarque cependant une parenté sémantique tout à fait frappante entre les dix premiers verbes, dont une majorité renvoie à un processus de structuration cognitive. Ceci suggère donc une co-configuration de *along* par un certain type de procès *x* et un certain type de repère *y* qui aboutit à la valeur que nous avons appelée *argumentative*.
- 39 Dans l'échantillon de 1000 occurrences extraites du BNC, ce type de valeur recouvre au moins 10 % du total<sup>20</sup>.
- 40 Une autre séquence typiquement argumentative est constituée du lemme *go* suivi de *along* *with*. Le tableau 4 montre les résultats d'une requête portant cette fois-ci sur les lemmes nominaux dans une fenêtre de trois mots à droite de cette construction<sup>21</sup>.

Tableau 4 : Résultats d'une requête portant sur les lemmes nominaux dans une fenêtre de trois mots à droite après la construction *go along with*...

Rank	Lemma	Log-Likelihood
1	wish	60.1536
2	plan	39.2626
3	idea	26.5394
4	proposal	21.7481

- 41 À nouveau, les probabilités sont, en termes purement quantitatives, peu significatives, mais la parenté sémantique est remarquable entre ces termes, qui renvoient autant à la conation qu'à la cognition.
- 42 Nous avons regroupé sous un seul chapeau les configurations argumentatives et subjectives de *along* car il est effectivement parfois délicat de les séparer : souvent, l'alignement d'un sujet avec un mouvement argumentatif implique un alignement subjectif avec le porteur de l'argument. On peut estimer que la différence se trouve dans la nature de l'espace séquentiel défini par *y* qui, dans le cas des configurations subjectives, renvoie à une trajectoire subjective menant du projet, de l'intention ou du désir, vers la réalisation.
- 43 On peut citer à ce titre les exemples (13) et (14), qui renvoient tous deux à ce schéma d'alignement subjectif *go along with somebody* dans (13) et simplement *go along* dans (14), interprétable subjectivement comme *go along with her* ou spatialement, de manière moins immédiate, comme *go along to the south coast*, etc.

13) He was still not sure that he believed that story, but in her misery was prepared to go along with her. HGE 2991

14) 'I know, but my mother liked the south coast and London for holidays, and Dad just went along ; if it suited Mother it suited him.' H8J 1139

- 44 Le procès *come* renvoie de manière déictique à un déplacement en direction d'une source subjective. L'association de ce mouvement déictique et coorienté, aboutit typiquement à des valeurs de persuasion, de conciliation, etc. où un sujet tente de ramener un deuxième sujet vers sa trajectoire personnelle.

15) Now, Maire, enough of this whining. It's not a bit nice. Benny has said she was sorry. You know it was an accident. Come along now and be a big girl. CCM 498-502

16) Richard Up up this arm in. Put this little arm in. [...] No. Come along now. There we go. That's it. KDP 2600-2605

- 45 Quant au procès inessif *get*, dans *get along*, il construit l'espace séquentiel *y* définit par *along* comme un objectif à atteindre, aboutissant sur une continuité subjective et spatio-temporelle avec une trajectoire privilégiée (17), éventuellement de façon réciproque (18).

17) You can't get along without it. HR7 3536
18) 'Tessa's fine, we still get along together fine,' he said. K35 139

## 7. Remarques sur quelques expressions polylexicales comportant *along*<sup>22</sup>

- 46 L'exploration du corpus par le genre de requête que nous avons présenté permet de mettre à jour des constructions complexes, qui relèvent, dans le modèle que nous revendiquons, d'opérations de co-configuration entre marqueurs<sup>23</sup>. La suite *along with*, classée par Quirk et al (1985 : 669 ; 761) comme une préposition complexe ou quasi-coordonnante, se caractérise non seulement comme comitatif mais également comme l'indice d'un mouvement coorienté dans le cadre d'un projet partagé ou d'un phénomène commun. Cette caractérisation se vérifie dans une requête portant sur les lemmes nominaux dans une fenêtre de trois mots à droite de *along with* (tableau 5). Les collocations évoquent, pour la plupart, l'appartenance à un ensemble défini par une finalité partagée (*colleague*, *team-mate*, etc.).

Tableau 5 : Résultats d'une requête portant sur les collocations par lemme dans une fenêtre de trois mots à droite de la séquence *along with*

- 47 Les exemples (19) et (20) fournissent une illustration de cette valeur complexe, d'abord avec l'association de *host DNA* à l'intérieur d'un virus (19), puis de manière sans doute plus intuitive, avec l'association d'un membre de Manchester United *Moran* avec d'autres co-équipiers *David Platt* et *Peter Beardsley* (20).

19) As the virus DNA multiplies, the piece of host DNA multiplies along with it. ANX 1513
20) But Moran — along with ex-United men David Platt and Peter Beardsley — would disagree. HAE 252

- 48 Un cas intéressant concerne la séquence *all along*. Ce bloc, qui associe un marqueur de fermeture, ou de bornage *all*, et le marqueur *along*, qui construit une séquence non bornée, est interprété comme un intensif. Il s'agit d'un procédé bien connu que l'on retrouve en anglais dans l'association de la forme BE + ING et *always* ou encore de *each* et *every*. L'interprétation intensive bloque les valeurs spatio-temporelles, pour ne laisser que des valeurs subjectives ou argumentatives. Les collocations de *all along*, indiquées dans le tableau 6, impliquent un certain nombre de termes qu'on pourrait qualifier d'épistémiques<sup>24</sup>.

Tableau 6 : Résultats d'une requête portant sur les lemmes dans une fenêtre de trois mots à gauche de la séquence *all along*

Rank	Lemma	Log-Likelihood
1	have	422.9809
2	know	404.4067
3	right	112.2202
4	be	85.1879
5	say	49.1879
6	suspect	35.2927
7	intend	32.7974
8	i	32.3932
9	there	32.3454
10	maintain	24.1422

- 49 Les exemples en contexte évoquent typiquement la révélation de discrédits subjectives (21) et (22).

21) Everyone except Robson knew all along he was world class ACN 555
22) She wanted to be alone for a minute, away from Hepzibah's kindness and Albert's triumphant look. Of course, he'd been right all along ! EFJ 3105-3106

- 50 Un dernier cas de figure, problématique, et qu'il serait intéressant d'explorer davantage, concerne la suite *be along*, telle qu'on la voit dans les exemples (23) et (24).

23) My sisters have been along to a couple of events BMM 2262
24) 'I told you, she'll be along in a minute.' FP0 1831

- 51 Cette configuration pose problème dans la mesure où on ne s'attend pas à l'association du dynamisme de *along* avec le statisme de *be*<sup>25</sup>. La clé, qui demande vérification, est sans doute à chercher dans le fait que cette configuration semble ne se trouver qu'à l'accompli (*have been along* dans le (24)) ou à une forme modale de projection vers l'avenir (*'ll be along* dans le (25))<sup>26</sup>. Dans le cas de l'accompli, *be* paraît équivalent de *come* ou *go*, situation analogue à celle de *be to* dans *Mr Lewis had been to Spain A1B 342*, par exemple. Dans le cas du renvoi à l'avenir, nous faisons l'hypothèse que c'est le passage dynamique de la situation repère vers la situation visée qui permet de reconstruire l'espace séquentiel déterminé par *along*.

## 8. Discussion et conclusion

- 52 Au terme de l'introduction consacrée aux possibilités de représentativité d'un corpus, nous avons choisi de reformuler la question, en quelque sorte, en nous interrogeant sur la question de la représentativité du corpus selon le cadre théorique on l'on se situe et les objectifs que l'on se fixe. Les rapports entre les méthodologies de la TOPÉ et celles de la linguistique de corpus peuvent sembler bien lointains. Les linguistes de corpus auraient parfois raison de reprocher aux linguistes énonciativistes leur tendance au « cherry-picking » : l'utilisation du corpus comme vivier d'exemples seulement, où l'on peut puiser à sa guise ceux qui servent le mieux la thèse défendue. Et les énonciativistes auraient parfois raison de reprocher aux linguistes de corpus une trop grande dépendance à l'égard des données quantitatives et une tendance concomitante à privilégier l'outillage mobilisé (taille du corpus, visualisations sophistiquées) sur la théorisation.
- 53 Dans le présent article nous avons voulu défendre la modélisation énonciative du langage prônée par Culioli, tout en nous appuyant sur les résultats de requêtes ciblées sur corpus. La TOPÉ insiste sur la nécessaire prise en compte du contexte, dans la description des valeurs d'un marqueur, et sur la façon dont les opérations auxquelles renvoie une forme schématique sont paramétrées par les opérations d'autres marqueurs contextuellement présents. Lorsqu'on travaille sur des corpus suffisamment étendus et suffisamment bien annotés (lemmes et parties du discours, notamment) l'utilisation raisonnée de méthodologies de corpus permet, à notre avis, de *quantifier* l'apport *qualitatif* du contexte, et de mettre ainsi à l'épreuve l'heuristique de jugement du linguiste énonciativiste<sup>27</sup>.
- 54 Nous avons voulu démontrer la pertinence de cette démarche par une analyse du marqueur *along* qui confronte tour à tour les analyses qualitatives menées en contexte aux résultats d'analyses quantitatives sur corpus. Nous avons ainsi mis en lumière la polyvalence du marqueur *along*, en posant que l'espace séquentiel non borné, dont *along* indique le statut de repère, peut, selon des éléments contextuellement localisables, porter sur les différents paramètres de la situation d'énonciation que sont l'espace, le temps, le sujet ou encore le discours (l'argumentation)<sup>28</sup>. Nous sommes loin d'avoir épuisé l'exploration du marqueur, qui pourrait par exemple étendre l'analyse à la distribution des types de *along* en fonction des registres, ou encore à l'étude des possibilités de composition qui aboutissent à des lexèmes tels que *sing-along*, *roll-along*, *jog-along*...<sup>29</sup>, tâche que nous envisageons d'entreprendre dans le cadre d'une prochaine étude.

---

## BIBLIOGRAPHIE

Ashino, Fumitake, Jean-Jacques Franckel & Denis Paillard. 2017. Prépositions et rection verbale : étude des prépositions « avec », « contre », « en », « par », « parmi », « pour. » GRAMM-R 39. Frankfurt am Main : Peter Lang.

- Barcelona, Antonio. 2003. Introduction The cognitive theory of metaphor and metonymy. In Antonio Barcelona (ed.), *Metaphor and Metonymy at the Crossroads*, 1–28. Berlin, Boston : De Gruyter.
- Biber, Douglas (1993). “Representativeness in Corpus Design.” In : *Literary and Linguistic Computing* 8.4, pp. 241–257.
- Col, Gilles. 2017. Construction du sens : un modèle instructionnel pour la sémantique (Linguistic Insights Volume 224). Berne ; New York : Peter Lang.
- Dufaye, Lionel. 2014. Over : A Comparative Study of Polysemy in Enunciative and Cognitive Approaches. *Faits de langues*. <https://hal-upec-upem.archives-ouvertes.fr/hal-01689395>.
- Franckel, Jean-Jacques & Denis Paillard. 2007. *Grammaire des prépositions*. Paris : Ophrys.
- Habert, Benoît. (2010). Des corpus représentatifs : de quoi, pourquoi, comment ? In M. Bilger (ed.), *Linguistique sur corpus – études et réflexions*, Perpignan : Presses Universitaires de Perpignan.
- Hoffmann, Sebastian, Stefan Evert, Nicholas Smith & David Lee. 2008. Corpus linguistics with BNCweb : a practical guide. *English Corpus Linguistics* v. 6. Frankfurt am Main : Peter Lang.
- Langacker, Ronald W. 2010. Reflections on the Functional Characterization of Spatial Prepositions. *Corela* HS-7. doi :10.4000/corela.999. <http://journals.openedition.org/corela/999> (10 February, 2019).
- Quirk, Randolph, Sidney Greenbaum, Geoffrey N Leech & Jan Svartik. 1985. *A Comprehensive grammar of the English language*. (Ed.) Randolph Quirk. India : Longman / Dorling Kindersley.
- Ranger, Graham. 2014. « Fire away » Suggestions pour une caractérisation énonciative de la particule adverbiale anglaise AWAY. In J.-M. Merle (ed.), *Faits de langues. Prépositions et aspectualité*. Bern : Peter Lang. pp. 147–158.
- Ranger, Graham. 2018. *Discourse markers : an enunciative approach*. New York, NY : Springer Berlin Heidelberg.
- Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collocations : Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–243. doi :10.1075/ijcl.8.2.03ste.
- Tyler, Andrea & Vyvyan Evans. 2007. *The semantics of English prepositions : spatial scenes, embodied meaning and cognition*. Digitally printed version (with corr.). Cambridge [u.a] : Cambridge Univ. Press.
- Vandeloise, Claude. 1987. *Les Prépositions en français*. Paris : Seuil.
- Van der Gucht, Fieke, Klaas Willems & Ludovic De Cuyper. 2007. The iconicity of embodied meaning. Polysemy of spatial prepositions in the cognitive framework. *Language Sciences* 29(6). 733–754. doi :10.1016/j.langsci.2006.12.027.

## NOTES

1. Cf. par exemple Biber (1993) ou Habert (2010).
2. Par exemple, dans Culioli (2002: 131 et sq).
3. On pourrait également considérer que la démarche favorise l'étude de cas d'usage exceptionnels, car le récit littéraire se distingue à la fois par le respect et, ponctuellement, le non-respect, des formes canoniques du genre.

4. « Dès qu'on travaille de façon fine, on se heurte à cette question de la collecte des données fiables. Une fois de plus, surgit la nécessité du travail théorique: il s'agit cette fois de construire une théorie des observables » (Culioli 1991: 17).
5. Cf. Hoffmann et al. (2008). L'interface de requête se trouve au lien suivant: <http://bncweb.lancs.ac.uk/>
6. On pourra mentionner à ce titre les travaux fondateurs de Lakoff (1987) ou de Vandeloise (1986), ainsi que les travaux plus récents de Barcelona (2003), Tyler & Evans (2007), Van der Grucht et al (2007), Langacker (2010) ou encore la monographie innovatrice de Col (2017).
7. Nous concédons que d'autres traductions, éventuellement meilleures, sont possibles; il s'agit ici simplement de présenter le problème par le biais particulier de la traduction.
8. La requête est relativement simple sur l'interface BNCweb: on recherche la forme (*along* ou *alongside*) puis on calcule les collocations, en incluant les annotations de lemmes, on choisit ensuite « collocations on lemma », en limitant la fenêtre et en précisant la classe de lemme VERB ou SUBST. Le calcul se fait par défaut sur la probabilité logarithmique.
9. La probabilité logarithmique n'est sans doute pas l'indice le plus indiqué pour une comparaison de ce type, mais nous avons pu constater que d'autres mesures statistiques aboutissent à des constats largement similaires.
10. Il est difficile de savoir comment formuler ce constat; on pourra considérer que tout objet concret possède une longitudinalité, dès lors qu'il ne s'agit ni de polygones réguliers, ni de cercles. Cela dit, on admettra facilement que la longitudinalité, ou la propriété /LONGUEUR/ paraît plus saillante pour une route, un couloir ou une rivière, que pour un parking, une salle ou un lac, par exemple. Noter que les propriétés s'écrivent par convention entre barres obliques.
11. Cf. Culioli 1991: 76 et sq. pour la symbolique formelle.
12. Les références alphanumériques sont celles du BNC: l'identifiant du texte suivi de l'identifiant du segment.
13. Cette formule fait usage de  $\exists$ , l'image-miroir de  $\underline{\exists}$ , qui indique une relation où le repère est à gauche, et le repéré à droite.
14. On trouvera une schématisation similaire dans Ashino, Franckel et Paillard (2017) par exemple. Le lecteur consultera également avec profit les travaux antérieurs de Franckel et Paillard (2007).
15. Cf. Ranger (2014) pour une étude de *away* qui adopte le même type de modélisation.
16. La schématisation se heurte comme toujours à la difficulté qui consiste à représenter par des figures statiques ce qui est par nature dynamique.
17. Voici l'ensemble des catégories employées avec les fréquences relevées: préposition (spatiale: 473, argumentative: 44, temporelle: 14), particule (spatiale: 162, argumentative: 56, temporelle: 32), locution *along with*: 202, unclear: 13, other: 4. Le chiffre de 63,5% regroupe les prépositions et les particules à référence spatiale (635 occurrences sur 1000).
18. Cf. Dufaye (2014) pour une comparaison des approches cognitive et énonciativiste de la polysémie portant sur le marqueur *over*.
19. L'exemple suivant, cité par le Oxford English Dictionary, suggère que cet emploi temporel n'est pas une innovation récente: *Wuldres scima, æþele ymb æþelne, ondlonge niht scan scirwered Guthlac B 1287 (A ray of glory, noble around the noble one, all night long, shone clearly).*
20. Nous précisons au moins car la prochaine configuration, la locution *along with*, est souvent argumentative, elle aussi.
21. La suite *go along with* apparaît au total 483 fois dans le BNC. Le tableau ne comporte que quatre items, puisque la requête sur collocations était formulée avec une fréquence d'occurrence minimale imposée de cinq occurrences. En baissant celle-ci à trois occurrences, on retrouve *charade*, *view*, *theory* ou *system*, entre autres.
22. Par « expressions polylexicales » nous entendons l'équivalent de l'anglais « multi-word expressions ».



23. L'approche peut présenter des similitudes avec la grammaire des constructions, à ce détail près que dans le cadre qui est le nôtre, la valeur d'une construction de ce type reste globalement calculable à partir des opérations dont ses éléments constitutifs sont la trace.
24. La collocation « i », Rang 8, est le pronom personnel, transcrit en minuscule puisque lemmatisé.
25. Nous écartons de cette discussion des exemples tels que *The journey to Reykjavik [...] is along a very reasonable road* A6T 173 ou *The practice is along the following lines*. B07 948 qui se ramènent aux cas déjà étudiés.
26. La formule est maladroite: nous pensons à *will / would be along*, et d'autres cas similaires.
27. C'est une approche que nous défendons depuis plusieurs années, le plus récemment dans notre ouvrage consacré aux marqueurs de discours en anglais (Ranger 2018). Elle partage un certain nombre de points communs avec l'approche des « collostructions » décrite en Stefanowitsch & Gries (2003).
28. On se contente souvent de mentionner les paramètres S et T de la situation d'énonciation, sans considérer le paramètre E, qui renvoie au rapport entre un énoncé donné et le discours avoisinant. Or, celui-ci est d'une importance capitale dès lors qu'on se penche sur des phénomènes d'argumentation, de connectivité inter-énoncés, etc.
29. Les textes qui composent le BNC sont effectivement indexés selon leur « text-type », ce qui permet de revenir vers le corpus spécialisé dont on peut étudier les spécificités en comparaison à un corpus multi-genre.

## RÉSUMÉS

La représentativité d'un corpus est largement tributaire du rapport entre le corpus et la langue, ou variété de langue, dont le corpus représente un échantillon. Ce rapport ne peut s'évaluer de manière satisfaisante, car la langue cible ne peut être saisie que via un ensemble fini d'occurrences, en d'autres mots, un corpus. Nous préconisons par conséquent d'évaluer la représentativité selon les objectifs spécifiques que l'on se fixe. Dans le cadre énonciatif, les marqueurs d'une langue sont décrits en termes d'un invariant, ou forme schématique (FS), qui est configurée par les opérations des marqueurs associés, pour aboutir à des valeurs situées en contexte. L'élaboration de la FS se fait sur la base d'exemples authentiques, étudiés en contexte et soumis par le linguiste aux manipulations, jugements d'acceptabilité, etc. avec la part d'intuition que ce processus implique. Cet article défend l'utilisation des méthodologies de la linguistique de corpus dans un cadre énonciatif, afin de vérifier et d'orienter la modélisation théorique du langage. Cette méthode sera exposée par une étude de cas du marqueur "along" qui s'appuie sur des données issues du British National Corpus, mettant en évidence différentes valeurs contextuelles du marqueur, dérivables d'une FS, sur la base de mesures d'associativité.

The representativeness of a corpus is a function of the relationship between the corpus and a target language or language variety the corpus is intended to sample. However, the second term of this relationship can only be apprehended via a finite set of language occurrences, i.e. a corpus. For this reason we prefer to consider representativeness with respect to one's specific research goals. In the enunciative approach to language, markers are described in terms of an invariant schematic form, configured by the pressure of surrounding operations into contextually-situated shapes. The elaboration of a schematic form is carried out on the evidence

of authentic examples studied in context and submitted to manipulations and judgements of acceptability on the basis of the linguist's intuition, essentially. This paper argues that corpus methodologies can provide invaluable quantitative support for these intuitions and, from there, for the theoretical constructions of enunciative approaches to language. We argue the case for such an approach with a study of the marker “along,” using data from the British National Corpus. Collocational affinities of “along” will allow us progressively to identify four particularly significant contextual configurations, yielding spatial, temporal, subjective and argumentative values for the marker.

## INDEX

**Mots-clés** : énonciation, Théorie des Opérations Prédicatives et Énonciatives, linguistique de corpus, préposition, particule, British National Corpus

**Keywords** : enunciation, Theory of Enunciative and Predicative Operations, corpus linguistics, preposition, particle, British National Corpus

## AUTEUR

**GRAHAM RANGER**

EA 4277 ICTT, Avignon Université