

Défense de la linguistique en tant que science empirique à la lumière de la conception scientifique de la biologie chez Mario Bunge

Dorota Zielińska

Suivez les preuves là où elles vous mènent et remettez tout en question.

Neil deGrasse Tyson

RÉSUMÉ — Bien que peu de linguistes adoptent actuellement le paradigme empirique, il y a une demande croissante pour le développement d'outils d'étude du langage similaires à ceux des sciences exactes. Cette tendance peut être observée même dans les principales revues linguistiques, telles que le *Journal of Pragmatics*, comme l'illustre Xiang (2017). Aujourd'hui, cependant, les linguistes qui adaptent les méthodologies des sciences plus avancées sont isolés de la communauté linguistique. En effet, la majorité des linguistes des départements de philologie et de philosophie restent convaincus que l'objet de leurs études est fondamentalement différent de ceux étudiés par les physiciens. Par conséquent, ils soutiennent que la méthodologie en linguistique ne s'apparente pas à celle utilisée dans les sciences empiriques. Dès lors, la linguistique est souvent perçue comme une science interprétative plutôt qu'explicative, et l'évaluation de la recherche en linguistique se fonde alors sur le consensus au sein de la communauté scientifique plutôt que sur des tests empiriques.

ABSTRACT — Although few linguists currently embrace the empirical paradigm, there are increasing calls for the development of tools for studying language that resemble those in exact sciences. This trend can be observed even in top mainstream linguistic journals, such as the *Journal of Pragmatics*, as exemplified by Xiang (2017). Today, however, linguists who adapt the methodologies from more advanced sciences face isolation from the mainstream linguistic community. This is because the majority of linguists in philological and philosophical departments remain convinced that the object of their studies is fundamentally different from those studied by physicists. Therefore, they argue that linguistic methodology cannot resemble that used in empirical sciences. As a result, linguistics is often seen as requiring interpretation rather than an explanation, and evaluation of linguistic research is based on acceptance within the scholarly community rather than empirical testing.

Dans cet article, nous analyserons de manière critique plusieurs arguments couramment utilisés pour soutenir l'idée fautive selon laquelle la méthodologie en linguistique ne peut pas ressembler à celle utilisée dans les sciences empiriques reconnues et nous montrerons qu'ils ne résistent pas à un examen approfondi. Pour ce faire, nous nous inspirerons de la célèbre défense de la biologie en tant que science empirique, formulée par Mario Bunge dans les années 1960, à une époque où de nombreux biologistes s'opposaient farouchement à un traitement scientifique de leur discipline.

1] Histoire des lois linguistiques

Il y a un demi-siècle, les biologistes traditionnels s'opposaient encore fermement à l'introduction de la méthode scientifique dans leur discipline. Certains groupes de biologistes ont avancé une pléthore d'arguments pour s'opposer à ce que la biologie soit traitée de la même manière que les physiciens abordent les phénomènes physiques. Ces arguments allaient de l'objection selon laquelle les organismes vivants ne peuvent être étudiés de la même manière que la matière inanimée, à la mise en évidence du rôle particulier de la méthode comparative en biologie. Aujourd'hui, aucun biologiste ne remet en question la valeur de la biologie moléculaire, de la biotechnologie, de la génétique ou de l'épigénétique – des disciplines fermement ancrées dans le paradigme empirique – même si la méthode comparative pour classer les espèces végétales et animales est toujours de rigueur.

Une dispute similaire avait également eu lieu entre les psychologues et les sociologues avant que nombre d'entre eux n'adoptent le paradigme empirique. Et qu'en est-il de la linguistique? La linguistique et les sciences empiriques doivent-elles appartenir à deux cultures distinctes dont les méthodes de recherche et les critères d'évaluation sont incompatibles, comme le suggère Snow (2001)? Alors qu'un nombre croissant de linguistes non conventionnels, qui adaptent la méthode scientifique à l'étude du langage, répondent «non» à cette question, les professeurs de linguistique les plus en vue continuent de défendre l'idée qu'il existe un fossé entre les sciences et la linguistique. Pourquoi?

Selon Grzybek (2006), de nombreux linguistes contemporains s'opposent à considérer la linguistique au même titre que les sciences

considère que le langage naît dans une société et se développe grâce à la réaction de ses membres aux propriétés et aux exigences de leur environnement par le biais de mécanismes d'adaptation, une règle de grammaire, entendue comme la description d'une structure linguistique grammaticale, ne peut être considérée que comme une tendance probabiliste de la *parole située*. Étant donné que les lois probabilistes concernent des tendances qui, par définition, ne peuvent pas s'appliquer à des cas individuels, toutes les grammaires, qui reflètent simplement des tendances, non seulement fuient, mais elles ne peuvent faire autrement que de fuir. Les contre-exemples à ces lois (fuites) au niveau d'un cas unique (l'occurrence d'une chaîne de mots), non seulement ne réfutent pas ces lois statistiques, mais sont anticipés et peuvent être déterminés quantitativement.

En résumé, il est essentiel de noter que toutes les grammaires, considérées comme des descriptions concises des structures linguistiques grammaticales, non seulement ont des fuites, mais doivent avoir des fuites. En outre, des lois apparemment rudimentaires comme le décompte d'occurrences et l'analyse de leurs interrelations, telles que «plus de A que de B», «plus il y a de A, plus il y a de B ou moins il y a de B», permettent d'explorer de nouveaux territoires dans tous les domaines des sciences empiriques, qu'il s'agisse de la linguistique ou de la physique.

3] Conclusion

L'opposition des linguistes aux efforts d'adopter l'approche empirique en linguistique s'explique par l'histoire des échecs chez les linguistes traditionnels d'identifier des lois déterministes et descriptives qui régiraient la grammaire et la signification d'une langue. Les linguistes traditionnels souhaitaient découvrir des lois linguistiques descriptives semblables à celles que Kepler a dégagées à partir des données de Tycho Brahe sur le mouvement des planètes autour du soleil, sans se rendre compte que la grammaire d'une langue ne peut pas être condensée en de telles règles déterministes en raison de la nature même du langage.

Lorsque les tentatives de trouver des règles descriptives et déterministes dans les données linguistiques ont échoué, les linguistes traditionnels ont conclu à tort que le langage ne pouvait pas être étudié dans le cadre du paradigme empirique. Cette croyance a donné lieu à un certain nombre de mythes censés la corroborer,

dont certains ont été démontés dans cet article. Cependant, l'existence d'un groupe de linguistes, souvent des physiciens reconvertis, déjà engagé dans des recherches sur le langage dans le cadre du paradigme empirique, constitue peut-être l'argument le plus fort contre cette croyance. Koehler (2012) présente un aperçu de plus d'une centaine de lois linguistiques élaborées dans le cadre de ce paradigme. Dans le présent article, nous avons discuté de deux autres groupes de lois linguistiques conçues dans le cadre du paradigme empirique : l'un concerne l'ordre des adjectifs dans les phrases nominales (Adj+Adj+Nom), l'autre concerne l'ordre des propositions temporelles contrefactuelles.

En plus de réfuter les mythes usuels qui alimentent la croyance selon laquelle le langage ne peut être étudié dans le cadre d'un paradigme empirique, cet article présente également ce cadre de recherche. Pour ce faire, il faut tout d'abord considérer le langage comme un aspect d'un système matériel. Compte tenu de nos connaissances actuelles sur le cerveau, telles qu'exprimées par le modèle de Jeff Hawkins présenté dans son article « Computing Like the Brain: The Path to Machine Intelligence » (2013), il est raisonnable de supposer que le langage émerge et évolue au sein d'une société grâce aux mécanismes d'adaptation de ses membres qui réagissent aux propriétés et aux exigences de leur environnement, comme l'expliquent Altmann & Koehler (2007)²¹. En particulier, les solutions linguistiques efficaces (utilisation d'éléments fréquemment utilisés, plus courts, ou qui ressemblent à d'autres éléments déjà bien maîtrisés) sont conservées en mémoire, ce qui entraîne une autorégulation du système, sans que les locuteurs recherchent activement des solutions optimales²². Étant donné que le langage est de toute évidence un système auto-organisé et autorégulé, les mécanismes qui produisent le langage peuvent être étudiés de manière optimale dans le cadre du systématisme et de l'émergentisme tel que décrit par Bunge (2003).

En ce qui concerne l'étude de la signification à l'intérieur de ce cadre, il faut postuler le mécanisme qui permet à l'interlocuteur

[21] Cette idée a déjà été formulée par Altmann (1978), mais de manière plus générale.

[22] Cette hypothèse distingue de manière cruciale les approches néozipfiennes actuelles visant à décrire le mécanisme d'autorégulation du langage, du principe du moindre effort de Zipf qui postule que les locuteurs recherchent consciemment des solutions linguistiques optimales lorsqu'ils parlent.

d'évaluer la signification située, perçue dans une situation socio-naturelle spécifique (dans un acte de langage situé) à un stade donné du processus d'interprétation, mécanisme qui peut potentiellement servir d'intrant pour d'autres processus d'inférence. Avec le systémisme et l'émergentisme à l'esprit, la théorie des champs du langage de Zielińska (2007, 2019) stipule que la signification située est le résultat de la sélection par les interlocuteurs, dans leur champ expectatif, du ou des éléments correspondant le plus au contenu encodé des mots interprétés. Le champ expectatif représente les idées et les mots qui, il y a un instant, sont venus à l'esprit d'une personne comme étant susceptibles d'être exprimés par la suite au cours du processus d'interprétation. Le champ expectatif est établi en tenant compte de facteurs tels que a) les informations relatives à la situation sociale (acte de langage situé), y compris sa raison d'être et les contraintes environnementales, b) les informations appréhendées verbalement jusqu'à un certain moment lors d'un échange verbal, c) le contenu encodé des éléments en cours d'interprétation, et d) les associations formées au fur et à mesure de l'échange. L'ensemble de ces informations dépendent des connaissances, de l'expérience, des préjugés, des intérêts et de l'attention de l'interlocuteur²³. Chaque option du champ expectatif se voit attribuer une probabilité d'être intentionnelle. Les «significations situées efficaces», telles que définies ci-dessus, sont stockées dans la mémoire, construisant et régulant les langages idiosyncrasiques. Les tendances statistiques dans ces langages individuels correspondent à la *langue* – le langage de la communauté considérée comme une structure abstraite. Notons également que la signification d'un mot, ou d'une autre unité stable du langage, est stockée dans le cerveau non seulement avec sa forme, mais aussi avec les informations contextuelles auxquelles il a été corrélé, qu'elles soient verbales ou non verbales.

Les hypothèses décrites ci-dessus permettent d'étudier les aspects quantitatifs et qualitatifs du langage dans le cadre du paradigme empirique. Lors de l'examen des aspects quantitatifs du langage, vu comme un système dynamique, non linéaire, composé de paroles situées et soumis à des processus évolutifs, bon nombre

[23] Cela correspond aux modèles récents de perception visuelle. La perception visuelle est le résultat du traitement des stimuli reçus une demi-seconde plus tôt. Nous construisons des modèles à partir de notre expérience antérieure.

de ses caractéristiques ne peuvent être saisies quantitativement que sous la forme de tendances probabilistes des relations entre les éléments examinés.

Les lois quantitatives du langage, explorées pour la première fois par Zipf, comprennent des liens tels que la relation entre la longueur et le rang des mots, la complexité des constructions linguistiques et leur fréquence, ou la dépendance de la dynamique du flux d'informations en fonction de la taille de celui-ci. On a constaté que nombre de ces liens suivent des distributions de type loi de puissance, une caractéristique commune à d'autres systèmes auto-organisés. En outre, nous pouvons étudier quantitativement des corrélations entre, d'une part, des ensembles de mots corrélés et, d'autre part, des mots de ce contexte. En particulier, nous pouvons examiner les corrélations entre les contextes d'énonciation des mots, ce qui permet d'obtenir un aperçu des relations de signification qu'entretiennent ces mots. Par ailleurs, il convient de noter qu'il y a plus d'un siècle, Firth faisait remarquer que «vous connaissez un mot à sa compagnie», annonçant ainsi l'approche relationnelle de l'étude de la signification.

Une autre approche de l'étude du langage dans le cadre du paradigme empirique consiste à tester des hypothèses quantitatives découlant des théories qualitatives de la compréhension et du traitement du langage. Cela peut se faire en analysant des corpus linguistiques ou en observant les réponses physiques accompagnant des interactions verbales, telles que la mesure des temps de réaction, l'enregistrement des mouvements oculaires ou la détection de l'activité cérébrale. Bien que ces hypothèses soient souvent rudimentaires, par exemple «plus il y a de A, plus il y a de B», les résultats de ces tests permettent une meilleure compréhension du langage, par exemple en validant les hypothèses qualitatives formulées.

Dans l'ensemble, il est clair qu'en linguistique, comme pour d'autres disciplines empiriques, la recherche quantitative est possible et vient compléter les études qualitatives. Les résultats quantitatifs peuvent être utilisés pour affiner les concepts linguistiques formulés qualitativement à l'origine, pour émettre de nouvelles hypothèses sur la base de ces résultats quantitatifs, ou encore pour tester rigoureusement les hypothèses qualitatives mêmes. Cependant, émettre des hypothèses qualitatives pertinentes, telles que passer d'une conception du langage comme une structure auto-

nome à une conception du langage comme un système auto-organisé et autorégulé, opérationnaliser de manière appropriée les concepts ou exploiter les données quantitatives est une opération essentielle du processus de création des idées, y compris les idées quantifiables.

Par exemple, supposer que le langage est un élément d'un système matériel doté d'un mécanisme d'auto-organisation et d'auto-régulation permet de rendre compte des lois de Zipf, les élevant du rang de simples futilités à celui d'argument central en faveur du langage comme système auto-organisé et autorégulé. Par ailleurs, proposer un mécanisme qualitatif d'interprétation de la signification dans le champ expectatif, qui dissocie la signification encodée des mots de leurs significations sélectionnées (situées), nous a permis d'acquérir, entre autres, de nouvelles connaissances sur le langage.

Premièrement, ce mécanisme a permis d'expliquer l'émergence de nouvelles significations. Ceci est crucial pour élucider la signification des mots utilisés dans des situations spécifiques, en tenant compte de la compositionnalité de la signification, de l'auto-régulation de la signification dans des langages idiosyncrasiques et, au final, de l'auto-régulation de la signification dans un langage communautaire. Deuxièmement, ce mécanisme offre une meilleure explication du phénomène de la transmission de messages différents par une même phrase dans divers contextes, allant au-delà de ce que permet de faire la division traditionnelle de la phrase en commentaire et en thème (le NOUVEAU et le DONNÉ). Selon ce point de vue, tout élément d'une phrase peut contribuer à identifier dans le champ expectatif de l'interlocuteur le thème ou le commentaire non codés (ou les deux), ce qui permet à la phrase de sélectionner un nombre de messages beaucoup plus important que ce que la division traditionnelle d'une phrase en DONNÉ et en NOUVEAU permet de réaliser. Ce mécanisme permet également d'expliquer le fait qu'une même phrase structurellement et lexicalement non ambiguë, utilisée dans des contextes différents (avec une structure d'information différente), peut avoir des représentations différentes et, par conséquent, des valeurs de vérité différentes. La raison en est que le champ expectatif postulé par le mécanisme de catégorisation de la FTL évolue au cours du processus d'interprétation. Par conséquent, lorsque les éléments d'une phrase sont interprétés dans un ordre différent (ce qui est le cas lorsque différents éléments sont traités comme le DONNÉ), les interprétations finales de cette

phrase peuvent différer les unes des autres. Enfin, il a été démontré que la FTL peut inspirer la création de lois linguistiques quantitatives sémantiquement pertinentes.

Remerciements. Je tiens à exprimer ma profonde gratitude à un certain nombre de personnes dont l'expertise et l'intégrité ont joué un rôle important dans l'orientation de mes recherches. Au regretté Jacob May, fondateur de la pragmatique linguistique, fondateur et rédacteur en chef du *Journal of Pragmatics*, pour son inspiration, ainsi que pour avoir édité l'article introduisant une première version de la FTL (Zielińska 2007a), de même que pour l'ajout de ses propres commentaires. Le professeur May m'a mis en relation avec Alessandro Capone de l'université de Messine. Il m'a porté assistance en de si nombreuses occasions que je ne saurais toutes les énumérer. Je lui suis particulièrement reconnaissante de m'avoir invitée à faire partie du comité de rédaction de la collection *Pragmatics Philosophy and Psychology*. Également, je remercie Martin Benes de l'université Charles en Tchéquie et de l'Académie tchèque des sciences, qui m'a agréablement surpris en traduisant mon livre en tchèque (Zielińska 2007b) et qui a également intégré ma théorie du langage à ses propres recherches, ce qui m'a particulièrement encouragé à continuer lorsque j'en avais le plus besoin. Je suis également redevable à Hélène Włodarczyk et à André Włodarczyk, respectivement professeure émérite à l'université Charles de Gaulle et professeur émérite à la Sorbonne, pour l'inspiration qu'ils m'ont apportée grâce à leurs recherches linguistiques influencées par leurs vastes intérêts allant de la sémiologie de l'art rupestre préhistorique à la cybernétique en passant par la poésie japonaise, ainsi que pour avoir remarqué ma propre théorie du langage et m'avoir invitée à la présenter à la Sorbonne. Enfin, je tiens à exprimer ma gratitude à François Maurice, fondateur de la Société pour le progrès des métasciences ainsi que de la présente revue, où la pensée de Mario Bunge joue un rôle important, pour ses commentaires utiles et sa patience éditoriale, ainsi qu'à Teresa Grabinska, physicienne théoricienne et philosophe des sciences, pour ses encouragements et son inspiration continus. C'est en lisant ses livres que j'ai entendu parler pour la première fois des recherches de Mario Bunge.

Références

- Altmann, G. (1985), «Sprachtheorie und mathematische Modelle», *SAIS Arbeitsberichte aus dem Seminar für allgemeine und indogermanische Sprachwissenschaft*, Vol. 8, p. 1-13.
- Altmann G. & Koch W.A. (1998), *Systems: New Paradigms for the Human Sciences*, De Gruyter
- Altmann G. & Köhler R. (2007), «Chapter 2: Quantitative Linguistics: An Overview», in S. Kepsner & M. Reis (ed.), *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives*, De Gruyter.

- Bruza P. *et al.* (2009), «Extracting Spooky-Activation-at-a-Distance from Considerations of Entanglement», in P. Bruza, S. Sofge, W. Lawless, K. van Rijsbergen & M. Klush, *Quantum Interaction: Third International Symposium, QI 2009, Saarbrücken, Germany, March 25-27, 2009. Proceedings 3*, Springer, p. 71-83.
- Bunge M. (1973), *Method, Model and Matter*, Reidel.
- Bunge M. (1959), *Causality: The Place of the Causal Principles in Modern Science*, Harvard University Press.
- Bunge M. (2003), *Emergence and Convergence: Qualitative Novelty and the Unity of Knowledge*, University of Toronto Press.
- Chierchia G. & McConnell-Ginet S. (1990), *Meaning and Grammar: An Introduction to Semantics*, MIT Press.
- Dummett M. (1993), *The Seas of Language*. Clarendon Press.
- Durka P.J. (2003), *Wstęp do współczesnej statystyki*, Adamantan.
- Grzybek P. (2006), *Contributions to the Science of Text and Language, Word Length Studies and Related Issues*, Springer.
- Jary M. (2008), «The Relevance of Complement Choice: A Corpus Study of “believe”», *Lingua*, 118, p. 1-18.
- Köhler R. & Altmann G. (1996), «“Language Forces” and Synergetic Modelling of Language Phenomena», in P. Schmidt (ed.), *Glottometrika 15: Issues in General Linguistic Theory and the Theory of Word Length*, WVT.
- Köhler R. & Altmann G. (2005), «Aims and Methods of Quantitative Linguistics», in G. Altmann, V. Levickij & V. Perebyjnis (ed.), *Problemy kvantytatynoilinguistyky/Problems of Quantitative Linguistics*, Ruta, p. 12-41.
- Köhler R. (2012), *Quantitative Syntax Analysis*, De Gruyter.
- Köhler R. & Altmann G., «Introduction to Quantitative Linguistics» (private communication).
- Kroupa P. *et al.* (2022), «Asymmetrical Tidal Tails of Open Star Clusters: Stars Crossing Their Cluster’s Práh Challenge Newtonian Gravitation», *Monthly Notices of the Royal Astronomical Society*, 517(3), p. 3613-3639.
- Lakoff G. (1987), *Woman Fire and Dangerous Things*, Chicago University Press.
- McEnery T. & Wilson A. (2003), *Corpus Linguistics. An Introduction*, Edinburgh University Press.
- McEnery T., Xiao R. & Tono Y. (2006), *Corpus-Based Language Studies An Advanced Resource Book*, Routledge.
- Nelson D. & McEvoy C. (2007), «Is There Something Quantum-Like About the Human Mental Lexicon?», *Journal of Mathematical Psychology*, 53, p. 362-377.
- Sampson G. (2001), *Empirical Linguistics*, Continuum.
- Sapir E. (1921), *Language: An Introduction to the Study of Speech*, Harcourt Brace.
- Searl J.R. (1983), *Intentionality: An Essay in the Philosophy of Mind*, Cambridge University Press.
- Siegel E. (2022), «We exist. What can that fact teach us about the Universe?», *Starts With a Bang!*, <https://medium.com/starts-with-a-bang/we-exist-what-can-that-fact-teach-us-about-the-universe-fdae9463a996>
- Skousen R. (1989), *Analogical Modeling of Language*, Kluwer.

- Snow C.P. (2001) [1959], *The Two Cultures*, Cambridge University Press.
- Ulam S.M. (1991), *Adventures of a Mathematician*, University of California Press.
- Wheeler J.A. (1994), *At Home in the Universe*, American Institute of Physics.
- Wulff S. (2003), «A Multifactorial Corpus Analysis of Adjective Order in English», *International Journal of Corpus Linguistics* 8(2), p. 245-282.
- Xiang M. (2017), «Toward a Neo-Economy Principle in Pragmatics», *Journal of Pragmatics*, 107, p. 31-45.
- Zipf G.K. (1932), *Selected Studies of the Principle of Relative Frequency in Language*, Addison-Wesley.
- Zipf G.K. (1935), *The Psycho-Biology of Language*, Houghton-Mifflin.
- Zipf G.K. (1949), *Human Behavior and the Principle of Least Effort*, Addison-Wesley.
- Zielińska D. (1999), «The Selective Mode of Language Use: The Way Natural Language Adapted Itself to Describing the World Around Us», *Zeszyty Naukowe UJ Prace Językoznawcze*, 119, p. 173-176.
- Zielińska D. (2003), «On the Selective Mode of Language Use», *Biuletyn Polskiego Towarzystwa Językoznawczego*, LIX, p. 27-35.
- Zielińska D. (2007a), «The Selective Mode of Language Use and the Quantized Communicative Field», *Journal of Pragmatics*, 39, p. 813-830.
- Zielińska D. (2007b), *Proceduralny model języka: Językoznawstwo z pozycji teorii modeli nauk empirycznych*, Wydawnictwo Uniwersytetu Jagiellońskiego²⁴.
- Zielińska D. (2014), *Proceduralní model jazyka: Lingvistika z pohledu teorie modelu empirických ved*, Edice Qfwfq, https://oltk.upol.cz/fileadmin/userdata/FF/katedry/kol/publikace/publ_qfwfq/Zielińska-Proceduralni_model_jazyka.pdf
- Zielińska D. (2019), «The Field Model of Language and Free Enrichment», in A. Capone, M. Carapezza & F. Lo Piparo (ed), *Further Advances in Pragmatics and Philosophy: Part 2 Theories and Applications. Perspectives in Pragmatics, Philosophy & Psychology*, vol. 20, Springer, p. 239-249.
- Zielińska D. (2020a), «How does language work?», (manuscript), https://www.researchgate.net/publication/344239308_3_How_does_language_work
- Zielińska D. (2020b), Testing the Advocated Theory of Language. The Study of the Order of Adjectives in Noun Clauses and of the Order of Counterfactual Before, [academia.edu/84422633/4_Testing_the_advocated_theory_of_language](https://www.academia.edu/84422633/4_Testing_the_advocated_theory_of_language)

[24] Puisque Zielińska (2007b) est épuisé, je propose ici sa traduction tchèque disponible gratuitement au format PDF : D. Zielińska, *Proceduralní model jazyka Lingvistika z pohledu teorie modelu empirických ved*, 2014, https://oltk.upol.cz/fileadmin/userdata/FF/katedry/kol/publikace/publ_qfwfq/Zielińska-Proceduralni_model_jazyka.pdf