

Extension sémantique de la préposition *en train de* : une perspective quantitative sur l'émergence diachronique de la polysémie

Quentin FELTGEN

Université de Gand

Résumé

La forme *en train de*, qui permet l'expression de l'aspect progressif, apparaît sémantiquement plus diversifiée que cette seule nuance. Pour éclairer sa structure sémantique, nous adoptons ici une approche diachronique. Nous montrons d'abord, par une étude qualitative, que le sens progressif émerge au XIX^e siècle, par grammaticalisation à partir d'un sens modal antérieur. Au XX^e siècle cependant, la fréquence d'occurrence connaît un nouvel essor. Différents indices quantitatifs permettent de caractériser ces nouveaux emplois de la forme. Il en ressort que *en train de* apparaît comme polysémique, l'usage actuel s'étant développé suivant deux processus diachroniques distincts et cumulatifs.

Abstract

The linguistic form en train de, which is used to convey a progressive aspectual reading in French, appears semantically more complex than this use alone. To investigate its semantic structure, we follow a diachronic perspective. We first show, through a qualitative survey of examples, that the progressive meaning emerges within the 19th century, grammaticalizing from a former modal meaning. In the 20th century though, the token frequency rises again. Several quantitative insights allow a better characterization of this new array of uses for the form. Insofar as it has developed through two distinct and cumulative diachronic processes, en train de can therefore be construed as polysemic.

Introduction

« Tout ce que l'on a appris au siècle dernier est en train d'être un peu dépassé » (Saloff-Coste 2019). C'est possiblement vrai, même si l'on est en droit de s'interroger sur ce que cela signifie : en quoi le dépassement des connaissances est-il un processus ? En ce que les connaissances sont dépassées les unes après les autres, si bien que certaines seraient toujours valides, d'autres non ? En ce que la validité des connaissances s'effrite progressivement, jusqu'à ce qu'elles soient tout à fait obsolètes ou remplacées – les connaissances étant en somme, et dès leur germe, en décalage croissant avec une épistémè qui s'en éloigne ? Ou bien faut-il voir dans cet énoncé un signe d'imminence : certes, tout ce que l'on a appris n'est peut-être pas encore dépassé, mais ne saurait tarder de l'être, du fait, peut-être, d'une accélération dans la production des savoirs associée au siècle nouveau ? Une part de l'ambiguïté sémantique de cet énoncé tient sans doute à la préposition *en train de*, et à la valeur aspectuelle précise qu'elle lui imprime.

Pourtant, la situation paraît simple ; *en train de* exprimerait le progressif, soulignerait que le procès qu'elle introduit est en cours et pas encore à terme. Ainsi, Langacker (1987) écrit que « les verbes perfectifs sont ceux qui admettent la construction *en train* [...] ; les verbes imperfectifs [...] n'admettent pas cette construction ». La construction *en train de* constituerait donc un critère de démarcation ; cependant, d'autres auteurs lui associent un certain flou fonctionnel qui s'accommode mal de la netteté de cette valeur perfective. Ainsi, Besse (1970 : 73) s'interroge sur les gloses possibles de la construction pour en traduire l'idée à des apprenants de français langue étrangère. Pour lui, la difficulté vient de ce que, pour ce type de locution, le sens dérive avant tout des « interprétants syntagmatiques », car « moins un signe a de sèmes, plus son signifié est déterminé par les contextes. » Alternativement, on peut considérer que c'est au contraire parce que *en train de* présente de multiples sens, que son emploi peut s'accorder à un plus large éventail de contextes, ne serait-ce que pour la possibilité qu'il a de se combiner avec des verbes téliques et atéliques (Vetters 1989) : tant qu'on est en train de marcher (atélique), on marche, tant qu'on en est en train de partir (télique), on n'est pas encore parti.

Notre objectif est ici de rendre compte de la pluralité des valeurs aspectuelles associées à *en train de*. Privilégiant une approche quantitative de la question, nous présenterons dans une première partie les approches quantitatives de représentation du sens qui ont fait récemment surface avec l'avènement des *Large Language Models*. Cependant, nous argumenterons que ces outils ne permettent pas facilement de trancher la question de la polysémie proprement dite, c'est-à-dire ne permettent pas d'identifier si les unités sémantiques présentent une ou plusieurs valeurs sémantiques (pour larges et étoffées qu'elles puissent être). Pour sonder la structure sémantique de cette construction, nous présenterons une autre approche, fondée sur une perspective diachronique. Celle-ci part du

postulat suivant : différentes valeurs sémantiques font l'objet de développements diachroniques distincts ; en particulier, si l'extension sémantique se traduit bien par une augmentation de fréquence d'occurrence suivant le motif de la courbe en S (Kroch 1989, Blythe & Croft 2012), alors la détection de ces courbes en S doit permettre de retracer l'apparition de différentes valeurs sémantiques.

Dans une deuxième partie, nous proposons un état de l'art sur la construction *en train de*, en mettant en particulier l'accent sur le développement diachronique de sa valeur progressive, qui se révèle divisée entre, d'une part, l'hypothèse de l'émergence d'un sens aspectuel progressif à partir d'un sens modal, au tournant du XIX^e siècle, et d'autre part, l'hypothèse que le sens progressif existerait depuis une période bien antérieure. Or, il est crucial, pour notre approche, de pouvoir retracer précisément l'apparition de la valeur progressive associée à *en train de*, pour pouvoir l'associer de manière fiable à une augmentation de la fréquence d'occurrence. Pour clarifier ce point, nous proposons donc une étude qualitative basée sur l'analyse d'une sélection d'exemples, et qui nous permet de poser que le sens progressif paraît bien être le résultat, au XIX^e siècle, d'une grammaticalisation à partir d'un sens modal.

La troisième partie est consacrée à une étude quantitative de la construction à partir des données de corpus issues de la base textuelle Frantext (1998-2024). Nous appliquerons le modèle de la courbe en S à la fréquence d'occurrence, et monterons, d'une part, qu'une première courbe en S correspond bien au développement du sens modal, aujourd'hui largement disparu, et une seconde au développement du sens progressif. Surtout, nous identifions une troisième courbe en S, ultérieure au développement du sens progressif et correspondant selon notre hypothèse à une valeur sémantique supplémentaire, que nous tenterons d'identifier sur la base d'indices quantitatifs. L'existence de cette courbe en S nous semble ainsi confirmer la nature polysémique du fonctionnement actuel de cette construction.

1. Représentation du sens

La question de la polysémie des unités linguistiques relève, pour nous, plus d'un choix dans la description de ces unités que de leurs propriétés sémantiques. En effet, on peut chercher à donner un aperçu de la diversité des emplois d'une préposition – c'est ce qui motive l'article célèbre de Brugmann et Lakoff portant sur la préposition *over* (1988) – ou bien tenter de définir l'invariant sémantique qui caractérise ce que ses emplois ont en commun, par exemple pour la préposition *par* (Hamma 2005). Cependant, si la plupart des unités linguistiques, même les plus concrètes, présentent une variation de sens, il est clair que ces variations ne sont pas de la même amplitude d'une unité à l'autre ; par exemple, la préposition complexe *en partance de* n'a pas la même variabilité sémantique

que *en considération de*. Nous souscrivons ainsi pleinement à l'analyse en ce sens de Kleiber (2008), qui identifie ces deux positions extrêmes (celle du tout-poly-sémique et celle de l'invariant sémantique) et plaide en faveur d'une polysémie authentique, qui ne concernerait qu'un petit nombre d'unités linguistiques, et que des critères de nature empirique permettraient en principe de dégager. En ce sens, nous proposons ici d'adopter une vision radicalement quantitative de la question, en discutant d'abord des représentations géométriques du sens qui se sont développées à partir des années 1990 et qui connaissent aujourd'hui un nouvel essor avec l'avènement des *Large Language Models*. Nous verrons néanmoins que ces représentations, si elles permettent de systématiser la représentation du sens et de poser la question de la polysémie en des termes nouveaux et féconds, ne permettent pas de fournir un critère définitif pour rendre compte de la pluralité des sens. C'est pourquoi nous nous tournerons finalement vers une autre approche, toujours résolument quantitative, mais enracinée cette fois dans le développement diachronique des usages associés à une unité linguistique.

1.1. Opérationnalisation du sens dans les modèles quantitatifs

Récemment, de nombreuses techniques de traitement automatique des langues ont proposé des outils pour opérationnaliser la représentation du sens à partir d'objets mathématiques, appelés vecteurs, qui consistent en un jeu de coordonnées numériques dans un espace abstrait de dimension généralement très grande (plusieurs centaines de dimensions généralement). Ces modèles se répartissent en deux familles principales : les modèles *type-based* (un vecteur pour une unité/pour un mot ou « type ») et les modèles *token-based* (un vecteur pour un « token », c'est-à-dire pour chaque occurrence d'un mot donné dans un corpus). Par exemple, les techniques dites de *word embeddings* (Mikolov *et al.* 2013) permettent de construire un vecteur sémantique associé à un mot, cette représentation faisant l'objet d'une procédure d'apprentissage-machine à partir de vastes corpus (Wikipedia, Google Books) de manière à ce que la représentation vectorielle des mots capture au plus près leurs relations de co-occurrence. La représentation sémantique qui en découle est alors, par construction, monosémique.

Les approches dites *token-based* (Heylen *et al.* 2015), popularisées par le modèle BERT (Devlin *et al.* 2018), consistent quant à elles à enrichir cette première représentation monosémique en prenant en compte, pour chaque occurrence, l'apport de son contexte immédiat dans son interprétation sémantique. Là encore, la représentation en vecteurs est informée par une procédure d'apprentissage-machine visant non seulement à prédire une occurrence masquée étant donné son contexte, mais également à évaluer la cohérence inter-phrastique entre deux énoncés successifs. Dans cette représentation, chaque unité linguistique reçoit donc autant d'encodages (via la spécification des coordonnées du vecteur associé) qu'elle présente d'occurrences dans le

corpus. De manière inattendue, ces deux familles de modèles reproduisent ainsi les deux extrêmes mentionnés par Kleiber : une vision exclusivement monosémique d'une part, une vision où chaque contexte vient imprimer une nuance de sens distincte aux unités linguistiques d'autre part.

Ces techniques ont pourtant ouvert à une réflexion sur la polysémie, et ce dès l'émergence des premières représentations géométriques du sens, telles que proposées par exemple par Ploux & Victorri (1998). Ces auteurs construisent l'espace sémantique à partir de « cliques », c'est-à-dire des ensembles de mots tous synonymes les uns des autres. Ces cliques sont vues comme des unités sémantiques minimales, qui permettent une représentation du sens des mots plus granulaire que les approches associant à un mot une unique unité sémantique, et qui ne revient pas non plus à associer à chaque occurrence sa propre représentation sémantique. Dans une telle représentation, la polysémie est un phénomène omniprésent (la plupart des mots sont associés à une diversité de cliques sémantiques sous-jacentes), et gradient : les cliques se distribuant dans l'espace sémantique, ce formalisme donne lieu à des « cartes sémantiques », où les différentes cliques se répartissent entre différentes régions de l'espace sémantique abstrait qui permettent de passer progressivement d'un macro-sens à un autre. Or, ces macro-sens (qui pourraient justement correspondre à une analyse polysémique à un niveau moins granulaire, tel que le propose Kleiber) peuvent faire l'objet, dans le modèle, d'une caractérisation automatique via une étape supplémentaire dite de clustering (Ji & Ploux 2003). Le clustering consiste à regrouper les cliques sur la base de leur proximité relative, de manière à constituer des groupes de cliques proches les unes des autres. Cela permet d'identifier différentes régions sémantiques caractérisées par la proximité entre les cliques qui s'y rencontrent – et, par suite, de concevoir la polysémie à différents niveaux d'analyse.

Une approche similaire consiste à partir des matrices de co-occurrence entre mots : chaque mot est alors décrit par la coordonnée de ses combinaisons avec les autres mots, donc là encore par un vecteur de coordonnées binaires (0 ou 1), aussi grand qu'il y a de mots dans le vocabulaire choisi, et valant 1 si la combinaison se rencontre dans le corpus et 0 sinon. Cette approche a été en particulier utilisée pour représenter le sens des constructions dites schématiques, c'est-à-dire se composant d'une partie fixe et d'une partie libre, comme par exemple la préposition complexe *in the midst of* + {N} (Desagulier 2022), l'auxiliaire modal *may* + {V} (Hilpert 2016), ou la construction {V} + *the hell out of* (Perek 2016) – ou même encore la construction *en train de* + {V} qui nous préoccupe ici. Dans la terminologie de la grammaire des constructions, on appelle *type* tout mot pouvant rentrer dans l'argument libre de la construction. Le sens d'une telle unité complexe est alors associé à la projection, dans l'espace sémantique, du vecteur associé à chacun des types qui se combinent avec elle (ce vecteur étant donné par les coordonnées de co-occurrence de ce type avec les autres mots du corpus). Nous verrons là encore comment cette représentation donne lieu à différentes évaluations de la polysémie.

1.2. Polysémie et clustering

Quelle que soit la méthode considérée (*token-based*, cliques, matrices de co-occurrence), le passage d'une perspective très granulaire, qui sert de fondement à ces représentations, à une représentation faisant émerger des unités de sens moins fines, plus proches de nuances sémantiques générales associées à l'usage d'un mot donné, peut se faire par regroupement des unités sémantiques au moyen d'un algorithme de clustering hiérarchique, qui vise à regrouper itérativement les unités les plus proches les unes des autres (la distance étant définie via la représentation, c'est-à-dire via le jeu de coordonnées dans l'espace sémantique abstrait associé aux unités). À chaque itération de l'algorithme, une nouvelle unité sémantique se trouve regroupée avec une autre ou avec un groupe d'autres unités sémantiques (chaque groupe étant appelé un *cluster*), ou bien deux clusters sont regroupés l'un avec l'autre. Le résultat de cet algorithme est un arbre avec, à l'extrémité, l'ensemble des unités regroupées en un seul cluster (représentation monosémique), à l'autre extrémité, chaque unité considérée isolément (représentation radicalement polysémique). Entre les deux se trouvent toutes les nuances de polysémie possibles : représentation à deux sens (deux groupes d'unités sémantiques), trois, quatre, etc., jusqu'à une représentation entièrement éclatée entre unités isolées. Ces groupes émergent automatiquement de l'analyse quantitative et permettent une représentation polysémique du sens des mots avec le degré de granularité choisi.

On retrouve ainsi cette approche dans la représentation des sens de la préposition *over* à partir d'une représentation vectorielle *token-based* (Fonteyn 2021 : 21). Perek (2016 : 172) propose une analyse du sens associée à la construction {V} + *the hell out of* en se fondant successivement sur une répartition des types de la construction entre deux ou quatre clusters : deux clusters sont associés ainsi à une répartition abstrait/concret, quatre à une répartition verbes d'émotion ou de sentiment/actions abstraites/actions concrètes/actions concrètes sans contact direct. L'atlas sémantique (Ploux *et al.* 2010) repose également sur du clustering pour grouper les cliques entre elles et faire apparaître des niveaux de sens plus généraux. Dans tous les cas néanmoins, le nombre de clusters doit être spécifié et n'émerge pas naturellement de l'analyse – quoiqu'il existe des tentatives de rationalisation du choix du nombre de clusters, par exemple à l'aide de la méthode des *scree plots* (cf. Perek & Hilpert 2017). Ces approches quantitatives permettent donc de décrire la pluralité des sens, mais non pas de répondre à la question : y a-t-il ou non pluralité des sens ?

Une autre possibilité, notamment mise en œuvre dans le cas des constructions complexes, consiste à utiliser des cartes de niveaux (Hilpert 2016), une technique qui a notamment été utilisée pour comparer différentes prépositions complexes entre elles (Desagulier 2022). Cette approche s'appuie sur la méthode des matrices de co-occurrence décrite plus haut, mais a le mérite de prendre en

compte, outre les différents types, les fréquences de co-occurrence de ceux-ci à l'intérieur de la construction étudiée. Pour ce faire, on projette d'abord la représentation vectorielle des types dans un espace à deux dimensions, à des fins de représentation. En rapportant ensuite sur cette représentation les occurrences particulières de la construction trouvées dans le corpus de travail, on peut représenter la densité d'occupation de cette construction sous forme d'une carte de niveau, qui permet d'apprécier comment se répartissent les usages de la construction par rapport à cette carte sémantique.

En effectuant un plan de coupe de la carte de niveau à une « hauteur » de densité d'occupation donnée, on représente ainsi la répartition de la construction sur la carte sémantique comme un petit nombre de sommets représentant chacun une nuance sémantique. Cependant, le nombre de sens ainsi construits dépend bien évidemment de la hauteur de coupe ; par ailleurs, cette représentation sous forme de densité d'occupation dépend d'un paramètre libre, qui mesure l'étalement dans l'espace sémantique de chaque occurrence. Si cette méthode présente donc l'avantage d'enrichir la représentation avec des données de fréquence de co-occurrence, elle n'est pas fondamentalement différente dans son principe de l'approche consistant à regrouper les unités sémantiques à travers des clusters.

La polysémie, dans ces espaces sémantiques, constitue donc une caractéristique fondamentale et implicite du sens des unités. L'intérêt de ces représentations vient néanmoins de ce qu'elles éclairent quant à la structure du sens des unités, vis-à-vis d'un espace sémantique sous-jacent. Ces représentations opèrent ainsi un déplacement théorique : il ne s'agit pas de savoir si les unités sont monosémiques ou polysémiques (par construction elles sont toutes polysémiques), mais quelle est la structure du sens de ces unités. Ainsi, on a pu montrer que ces clusters, ou pics de densité, sont structurés autour d'un prototype (Montes & Geeraerts 2022), tout en admettant des emplois plus rares et plus marginaux. Plus crucialement pour nous, ce que ces méthodes laissent entrevoir (indirectement, car elles ne donnent accès à l'espace sémantique qu'à travers les réalisations linguistiques des unités lexicales), c'est qu'il existe une structure, une topologie en somme de l'espace sémantique : on peut ainsi le concevoir, de manière imagée, comme une succession de bassins et de vallées, séparés par des barrières traversées par d'étroits cols. Or cette image peut très bien servir de fondement à une vision de la polysémie : certains mots occupent un bassin, d'autres plusieurs, et les différents sens (les différents bassins) sont reliés entre eux par des cols qui permettent le passage de l'un à l'autre.

Si les méthodes quantitatives de représentation du sens donnent une assise empirique robuste à cette idée, et proposent une manière cohérente d'aborder la question, elles présentent trois écueils. Nous avons évoqué les deux premiers : d'une part, le paysage présentant une certaine « rugosité », il n'est pas clair de savoir à quelle échelle il est pertinent de se placer pour

définir les bassins sémantiques, c'est-à-dire quelle échelle approcherait au plus près les représentations cognitives des individus où la notion d'unité ou de pluralité des sens se pose avec le plus d'acuité. Nombre de clusters, hauteur du plan de coupe, etc. : ces méthodes montrent que la représentation change suivant l'échelle considérée, mais ne permettent pas de choisir cette échelle. D'autre part, ces méthodes ne permettent pas une représentation directe de l'espace sémantique, celui-ci n'apparaissant finalement que comme le reflet des représentations des occurrences individuelles, ou des types individuels – qui plus est, d'une manière généralement associée à une seule construction/unité lexicale, ou à un petit nombre d'entre elles.

Enfin, ces méthodes présentent, de notre point de vue, un écueil considérable : la fiabilité de la carte ne peut pas être évaluée. Pour l'exemple de *en train de*, sa fonction de marqueur aspectuel progressif le rend potentiellement compatible avec une grande variété de verbes (cette variété, on le verra, étant d'ailleurs objet de discussion). Or, la carte aura tendance à privilégier les groupements lexicaux (sur la base des contextes d'emploi dans lesquels ces verbes sont utilisés), et ignorer les nuances d'aspect ; par exemple, Hilpert (2016) souligne la proximité (lexicale donc) entre *laugh* et *smile* sur la carte qu'il établit, mais *en train de rire* et *en train de sourire* ne renvoient pas nécessairement à la même valeur aspectuelle. Il est d'ailleurs significatif que la représentation qui nous semble la plus conciliable avec l'étude de la polysémie soit donnée par les cartes tirées des cliques de synonymie de l'atlas sémantique, et non par celles s'appuyant sur les co-occurrences. Or les relations de synonymie doivent faire l'objet d'une appréciation lexicographique : elles ne sont pas dérivées de manière quantitative, et ne sont généralement pas disponibles pour les unités plus complexes, comme *en train de*.

1.3. La perspective diachronique

En partant de cette vision du sens des unités linguistiques comme un territoire sémantique qu'elles viendraient diversement occuper, nous souhaitons proposer une approche toute autre, fondée sur une perspective diachronique. Cette perspective repose sur cinq hypothèses, dont deux que nous avons déjà énoncées : 1) que l'espace sémantique peut être conceptualisé de manière « topologique », avec des bassins sémantiques séparés les uns des autres ; 2) que ces bassins constituent les différentes valeurs sémantiques pertinentes que viennent couvrir les usages d'un mot polysémique au sens de Kleiber. Les hypothèses suivantes se rapportent au changement linguistique.

La troisième hypothèse de notre approche est de considérer que, dans l'usage qui est fait d'un mot, les différentes valeurs sémantiques sont exploitées autant que possible par les locuteurs : autrement dit que l'usage « remplit » les domaines fonctionnels associés à l'énoncé. Cette hypothèse a notamment fait

l'objet d'une récente validation empirique (Feltgen 2024), montrant que l'amplitude du changement de fréquence des unités linguistiques dépend de la taille du domaine fonctionnel (ce que nous avons appelé ici un « bassin »), tandis que la dynamique d'augmentation de fréquence correspond à un remplissage homogène de ce bassin. La quatrième hypothèse, c'est que les phénomènes d'extension sémantique correspondent au passage d'un « col », ce qui conduit une unité linguistique à exprimer un nouveau domaine fonctionnel, autrement dit à remplir un nouveau bassin sémantique. Enfin, la cinquième hypothèse, c'est que l'augmentation de fréquence d'occurrence correspondant au remplissage d'un nouveau bassin sémantique obéit à une courbe en S, c'est-à-dire à un motif mathématique particulier (la sigmoïde) qu'il est possible de détecter à partir des données de corpus. Ces deux hypothèses sont en fait reliées de manière proche : si la barrière entre deux bassins est trop basse, leur remplissage se fera suivant un seul et même mouvement dynamique ; si au contraire la barrière est trop haute, seul un des deux bassins sera rempli ; pour observer la dynamique de remplissage d'un second bassin, il faut donc que le passage entre les deux bassins soit juste suffisant, et c'est précisément dans ces conditions que la courbe en S émerge comme motif de la dynamique (Feltgen *et al.* 2017). Cette approche a notamment fait l'objet d'une première confirmation empirique dans l'étude de la construction *en plein* + {N} (Feltgen 2022) : un premier développement diachronique obéissant à une courbe en S entraîne la valeur spatiale de centralité (*en pleine mer*, *en pleine Beauce*), un second développement diachronique plus tardif, obéissant à une nouvelle courbe en S d'augmentation de la fréquence d'occurrence, correspond au développement de la valeur aspectuelle d'inscription dans un processus (*en plein essor*).

Pour résumer, nous considérons que la polysémie, au sens de l'apparition d'une valeur sémantique à la fois nouvelle et en même temps liée à la précédente, correspond à un développement diachronique propre ; qu'une telle extension sémantique a une répercussion dans la fréquence d'occurrence du mot (elle augmente) ; qu'enfin cette augmentation se fait suivant un motif mathématique constant et empiriquement identifiable, la courbe en S. Autrement dit, les augmentations « en S » de la fréquence d'usage signalent le passage d'une unité linguistique entre deux domaines sémantiques faiblement connectés. Si le seul observable de fréquence ne renseigne en rien quant au contenu sémantique des domaines impliqués, il permet de sonder la topologie du territoire sémantique d'une manière qui est indépendante de toute représentation, et de détecter l'existence de régions quasi séparées – donc d'établir une situation de polysémie. Telle est la démarche que nous nous proposons d'adopter.

2. *En train de*, un développement diachronique complexe sur le plan sémantique

Pour illustrer cette interaction entre développement structural du sens et motifs de variation de la fréquence, notre choix s'est porté sur la construction *en train de*, et ce pour plusieurs raisons ; d'abord parce qu'elle est associée à un large nombre d'attestations, ce qui en facilite l'étude quantitative ; ensuite parce qu'elle a fait l'objet de nombreux travaux qui ont très largement contribué à en dessiner le riche paysage sémantique, en synchronie comme en diachronie. Elle présente par ailleurs, nous le verrons, un scénario historique de développement atypique et en cela tout à fait intéressant.

2.1. Statut et signification

2.1.1. Statut

Nous avons choisi dans cet article de présenter *en train de* comme une préposition complexe, à l'instar de la forme *en voie de* (De Mulder 2019). Du point de vue de la grammaire des constructions, on peut aussi bien la décrire comme la construction [*en train de* + {Vinf}], c'est-à-dire en explicitant l'argument libre, même si, historiquement, elle a pu se combiner avec des noms (1). Si cette construction est essentiellement associée à la périphrase verbale [copule + *en train de* + {Vinf}], elle peut occasionnellement se rencontrer apposée à un substantif (2), ou même isolée, comme cadratif (3).

- (1) puisque je suis en train de confidences, noble vicomte, je veux vous dire ce que je ne voudrais pas insérer dans une dépêche officielle, mais ce que m'ont inspiré quelques observations personnelles (F.-R. de Chateaubriand, *Mémoires d'outre-tombe*, t. 3, 1848)¹
- (2) Un soir, Armand, venu fortuitement de très bonne heure, trouva M. l'abbé Gondrand, directeur de la conscience de Mme de Langeais, établi dans un fauteuil au coin de la cheminée, comme un homme en train de digérer son dîner et les jolis péchés de sa pénitente. (H. de Balzac, *Histoire des Treize*, 1835)
- (3) Vendredi 28 septembre. En train de jouer après le dîner au billard avec Popelin, j'entends, dans le salon, la voix de Zeller arrivé dans la journée, lisant tout haut. (J. de Goncourt, *Journal : mémoires de la vie littéraire*, t. 3 : 1879-1890, 1890)

La construction [copule + *en train de* + {Vinf}] peut par ailleurs être vue comme une « fille » d'une construction plus schématique [copule + groupe prépositionnel], vaste famille parmi laquelle les constructions du type [copule + préposition complexe + {Vinf}] forment un petit groupe avec des membres tels que *sur le*

1. Sauf mention contraire, tous les exemples utilisés sont tirés de la base Frantext (ATILF 1998-2023).

point de, à deux doigts de, en voie de, en mesure de, en passe de, en état de, en position de, etc. Ces constructions prépositives présentent généralement une valeur aspectuelle ou modale et peuvent ainsi entrer en concurrence avec des constructions co-verbales (Merle 2008 : 53).

La construction prépositive [*en train de* + {Vinf}] se retrouve par ailleurs avec des verbes qui débordent l'inventaire des copules effectué par Lauwers & Tobback (2010) ; on la trouve notamment combinée dans le corpus frTenTen20 avec un large paradigme de verbes associés à la vision et au regard (*imaginer, apercevoir, montrer, se voir, voir, observer, surprendre, photographier, filmer, découvrir, croire, visualiser* etc.). Ce trait semble partiellement partagé avec les autres constructions de la famille [préposition complexe + {Vinf}], ce qui pourrait amener à privilégier le niveau de lecture périphrastique [{V = copule/vision} + {préposition complexe} + {Vinf}] pour rendre compte de cette famille de constructions, afin de souligner les contraintes et les spécificités liées à leur emploi.

On notera enfin que *en train de* n'admet plus guère la possibilité pour *train* d'être modifié (il a pu l'être par *bon, beau* et *grand*), mais permet d'intercaler un adverbe entre *train* et *de* (*seulement, alors, aussi, précisément, etc.*). À titre d'illustration, sur le corpus frTenTen20, la requête [word="en"] [word="train"] [tag="R.*"] [lemma="de"] [tag="V.*"] renvoie ainsi 625 occurrences, qui semblent à première vue recevables, dont le nombre doit être comparé aux 680,000 occurrences associées à la requête sans l'adverbe intermédiaire ([tag="R.*"]).

2.1.2. Signification

La signification de *en train de* a pu être établie par le biais d'une approche compositionnelle, ce qui semble d'autant plus pertinent que *en* présente un potentiel aspectuel intrinsèque (Leeman & Vaguer 2014). Bottineau (2014) propose ainsi l'analyse sémantique suivante :

[Le schéma sémantique] peut se gloser comme suit : le locuteur fait en sorte que l'allocutaire identifie au sujet (*être*) une propriété modificatrice transitoire (*en*) comprise comme un processus de modification en cours comprise entre une source et une cible (*train*) dont il faut remonter à la source (*de*) pour redéfinir la teneur événementielle (*infinitif*). Ce faisant, le locuteur conduit pas à pas un rattrapage cognitif, une sorte de geste mental de mise à jour : [...] il dirige activement le processus par lequel la conscience interprétante est focalisée sur le déroulement, puis ramenée à son origine, et enfin amenée à redéfinir lexicalement le procès dont le début a échappé à l'attention.

L'analyse sémantique la plus complète est probablement due à Franckel (1989), qui détaille divers usages de *en train de*. La définition qu'il en propose est néanmoins très tributaire de son cadre théorique :

être en train de articule, sous forme d'une discordance, l'ancrage de P notionnellement non structuré sur la classe des instants et la structuration notionnelle de P opérée hors du plan temporel (Franckel 1989 : 65)

Cependant, il l'étoffe de nombreux exemples et met à jour une grande palette d'emplois pour la forme *en train de*. Son approche est ainsi tout à fait compatible avec notre vision de la sémantique : on peut tenter de décrire le domaine sémantique global couvert par la forme à l'aide d'une définition d'ensemble, ou mettre à jour certaines valeurs sémantiques qui constituent et structurent ce domaine. Franckel propose en outre que ces différentes valeurs sémantiques se regroupent en deux grandes familles : les emplois téléonomiques de type « procès en cours » (exemple 4), et les emplois antitéléonomiques de type « mise en garde et inadvertance » (Franckel 1989 : 67), auxquels il rattache encore la prise de conscience (p. 74), et le rejet/la démarcation (p. 75) – que nous illustrons au travers des exemples (5), (6) et (7), respectivement. Le domaine est donc structuré autour de plusieurs usages distincts ; dans la terminologie de Franckel (1989 : 67), « être en train de correspond à un mixte complexe ».

- (4) Autour de lui, une dizaine de clients riaient grassement de ce petit bonhomme aux cheveux blancs qui frétillait comme un poisson en train de frirer.
(P. Darkanian, *Le rapport chinois*, 2021)
- (5) Vous êtes en train de perdre un million, vous vous rendez compte de ça ?
(J. Echenoz, *Cherokee*, 1983)
- (6) – Tu as déjà entendu des bruits, la nuit ?
Elle a ri :
– Évidemment ! Tu n'es pas en train de découvrir qu'il y a des bruits, la nuit, dans les maisons de campagne, Albert ?
(J. Chantreau, *Avant que naisse la forêt*, 2016)
- (7) Tu te demandes si tu n'as pas fait une monstrueuse connerie. – Non, je me demande si ce que je suis en train de faire est bien ou mal.
(M. Droit, *Le Retour*, 1964)

Cette pluralité d'emplois apparaît également chez Lachaux (2005), qui distingue jusqu'à neuf effets de sens distincts associés à l'usage de *en train de*. L'inventaire de ces effets de sens, allant de l'auto-justification à l'ironie, vise à montrer l'importance, dans l'usage associé à cette périphrase, d'une « prise en compte de l'interlocuteur par le locuteur » (2005 : 125). Surtout, elle insiste sur sa dimension modale, dans la mesure où « être en train de constitue un 'alibi' spatio-temporel, un faux calque de lien entre représentation linguistique et réel » (2005 : 134). Elle conclut ainsi que la valeur aspectuelle ne joue qu'un rôle secondaire par rapport à cette forte dimension pragmatique :

La valeur purement aspectuelle de la périphrase, lorsqu'elle apparaît, n'apparaît jamais seule, pour elle-même, mais fait l'objet d'un réinvestissement modal et communicationnel à visée persuasive de la part de l'énonciateur, sous couvert de délimitation spatio-temporelle.
(Lachaux 2005 : 138)

À l'inverse, Anscombe (2007) range *en train de* parmi les « indicateurs aspectuels de déroulement processif », aux côtés d'autres « auxiliaires prépositionnels » que sont *en cours de*, *en passe de* et *en voie de*. Il souligne les contraintes de compatibilité basées sur les *Aktionsarten* de Vendler (1957) : *en train de* se combine aisément avec les activités et les accomplissements, moins avec les achèvements et les états. Les contraintes dont il fait état paraissent trop restrictives, mais c'est probablement parce qu'*en train de* déborde ce sens processif qu'il peut se combiner avec des verbes qui s'accordent mal avec cette lecture. Dans la veine monosémique, on rangera également l'étude de Gosselin (2011). Il y décrit *en train de* comme servant à préciser la visée aspectuelle d'inaccompli d'un sous-procès, que ce sous-procès corresponde à la phase initiale, médiane ou finale d'un procès plus global. La périphrase permet ainsi de souligner l'exclusion des bornes du procès. L'auteur explique ainsi comment *en train de* peut se combiner avec des co-verbes marquant la catégorisation des phases du procès, comme *commencer à* ou *finir de*, puisque ces co-verbes sélectionnent une sous-phase du procès, tandis que *en train de* leur imprime une visée aspectuelle d'exclusion des bornes.

Do-Hurinville (2007), qui privilégie explicitement une perspective monosémique, propose le trait [+dynamique] comme invariant sémantique de *en train de* ; de là l'hypothèse que *en train de* peut se combiner avec tous les *Aktionsarten* qui partagent ce trait (c'est-à-dire tous sauf les états). Cependant les achèvements ne sont inclus qu'au prix d'une coercion, et se retrouvent alors « transformés en accomplissements ». Do-Hurinville note par ailleurs que des verbes statifs se rencontrent à l'occasion avec *en train de*, et explique cela soit comme un phénomène de coercion encore, le verbe d'état étant ré-interprété comme un procès, soit comme une possible valeur inchoative.

Cette hypothèse d'une fonction inchoative associée à certains des usages actuels de *en train de* est par ailleurs défendue par Leeman (2012) :

Ainsi, des étiquettes, pour la périphrase, telles que « aspect progressif », « aspect continuatif » ne sont-elles pas entièrement satisfaisantes, dans la mesure où elles masquent la possibilité, pour être *en train de*, de marquer le début, le point de départ, donc l'aspect inchoatif. La préposition n'est pas que de « centrage », elle implique une étendue bornée qui peut être saisie en son début aussi bien qu'en son centre. (Leeman 2012 : 135).

L'avantage de cette proposition est notamment d'apporter un éclairage quant à la combinaison de *en train de* avec « des verbes d'aspect ponctuel, qui, donc, ne supposent pas de 'déroulement' » (p. 136). Leeman discute en outre la valeur de « discordance » mise au jour par Franckel, et voit dans *en train de* l'expression d'une rupture entre la situation précédente et une situation nouvelle qui s'installe désormais.

La diversité des emplois de *en train de* est également mise en évidence à partir d'un corpus de français oral contemporain par De Wit & Patard (2013). Elles distinguent six usages de la locution, ainsi que deux marginaux (moins de 5% des

occurrences), l'emploi purement progressif restant largement majoritaire dans leur corpus (60 % des occurrences). La périphrase peut ainsi servir 1) à décrire le caractère « toujours en cours » de l'action, au présent, mais aussi 2) dans un contexte énonciatif où l'on relate une action passée, distancée de l'énonciation ; pour exprimer 3) l'inaccompli ou 4) une valeur habituelle ; 5) pour souligner le caractère limité dans le temps de la validité de l'énoncé ; et enfin 6) avec une connotation (inter-)subjective qui lui imprime un tour modal, dans la veine de Lachaux (2005). La dimension modale que peut ainsi présenter *en train de* invite par ailleurs ces autrices à y voir une expression de la contingence épistémique, c'est-à-dire que le procès introduit par *en train de* « ne fait pas partie de ce que le locuteur sait ou est en mesure de prévoir [du procès], [et il est] dès lors perçu comme d'une certaine façon inhabituel ». On retrouve ainsi l'idée précédente de « discordance ».

Il ressort de cet examen que, mis à part la fonction aspectuelle progressive couramment admise, il n'existe pas de consensus plein sur la sémantique de *en train de*, si ce n'est qu'elle semble dépasser la simple expression de l'aspect progressif. Faut-il y voir l'expression d'une valeur autre, toujours active, qu'il s'agisse de l'inchoatif, d'une coloration modale, d'une dimension épistémique ? Ou bien la coexistence de deux familles de sens plus ou moins séparées, comme le suggère Franckel (ce qui irait dans le sens d'un paysage sémantique que nous qualifierions de « polysémique » d'après la définition que nous en avons donnée) ? Pour y répondre, nous allons désormais nous tourner vers le processus historique de constitution du sens de la construction.

2.2. Développement diachronique

2.2.1. L'approche lexicographique

Le développement diachronique de la forme *en train de*, prise dans un faisceau de formes avec *train* et *en train* avec lesquelles elle partage certaines de leurs mutations sémantiques, a fait l'objet d'une étude pionnière par Gougenheim dans sa thèse consacrée aux périphrases verbales (1929), où il retrace les attestations métalinguistiques de *en train de* chez les grammairiens et dans les dictionnaires. En dépit des biais parfois majeurs dont ces ouvrages font parfois preuve, cette approche lexicographique est intéressante, car elle permet de donner une interprétation sémantique de la construction contemporaine de son usage, et non pas distordue par la connaissance de son sens actuel, qui tend à projeter une lecture progressive sur les occurrences antérieures. Par exemple, on peut considérer l'occurrence de *en train de travailler* (8) comme une manifestation sûre du progressif, étant analogue à des emplois contemporains où ce serait la seule lecture possible ; mais les occurrences (9) et (10) montrent que *en train de travailler* pouvait présenter à la même époque une lecture modale d'humeur/disposition. En (8), on comprend ainsi que Flaubert, se sentant incliné à écrire (sens modal), préfère ne pas perdre

la journée du lendemain en un voyage à Rouen pour se consacrer à l'écriture ; ce n'est pas parce qu'il est actuellement occupé à travailler (sens aspectuel), qu'il ne peut pas venir le lendemain (ce qui introduirait un curieux décalage). On voit par cette illustration comment la lecture sémantique des occurrences est biaisée par la familiarité avec les emplois contemporains ; raison pour laquelle la perspective lexicographique permet de déjouer ce biais en restaurant certaines lectures sémantiques disponibles pour les locuteurs de l'époque.

- (8) Mon bibi, je n'irai pas demain dîner chez toi, parce que je suis en train de travailler, et que ces voyages à Rouen me font perdre presque une journée entière. (G. Flaubert, *Correspondance* : supplément (1864-1865), 1865)
- (9) – révassé. – pas en train de travailler. – écrit ceci et couché. (J. Barbey D'aureville, *Memorandum* (Premier), 1838)
- (10) J'étais si en train de travailler que je n'ai pu, en faisant l'*Abîme*, m'empêcher d'écrire en six jours une petite comédie en un acte, *la Grand'mère*, que je vous lirai. (V. Hugo, *Correspondance*, t. 2, 1849-1866 [1865], 1866)

Nous ne reprendrons pas en détail l'analyse, fouillée et très riche, de Gougenheim. Signalons simplement que *en train* apparaît dans le *Dictionnaire universel* de Furetière de 1690, avec le sens « adverbe qui se dit de la disposition à faire ou à continuer quelque chose », c'est-à-dire un sens essentiellement modal, et l'exemple : « Allons jusque là tandis que nous sommes en train de marcher. » (Gougenheim 1929 : 62). Le dictionnaire de l'académie n'enregistre *en train de* Vinf qu'à partir de l'édition de 1740, qui mentionne : « Être en train de jouer, de courir, etc. Être en humeur de jouer, de courir ; jouer, courir actuellement. Il est en train de se ruiner, Il mène une vie propre à le ruiner. », exemples auxquels l'édition suivante (1798) ajoute : « Il n'est pas en train de rire, Il n'est pas disposé à rire. ». Se trouve donc suggérée, en marge du sens modal, une valeur progressive (*jouer, courir actuellement*). Les éditions ultérieures (1835, 1870, 1935) n'apportent aucun nouvel élément. Le *Littré*, de son côté, ne fait que reprendre partiellement la définition de l'Académie (Gougenheim 1929 : 64).

Nous pouvons compléter cette analyse en remarquant que ni le *Dictionnaire de Trévoux* de 1738-1742 (qui reprend mot pour mot la définition du *Dictionnaire universel* et donne l'exemple « Nous sommes bien en train de nous divertir. »), ni le *Dictionnaire critique de la langue française* de 1787-1788 (qui ne donne *en train de* qu'en illustration d'un sens figuré d'« humeur » à l'entrée *train*, à travers l'exemple : « je ne suis point en train de rire »), ne mentionnent autre chose que la valeur modale. Il faut donc attendre la seconde moitié du XIX^e siècle pour rencontrer une attestation lexicographique explicite du sens processif, dans le *Grand dictionnaire universel du XIX^e siècle* (Larousse 1876 : 397) :

- Être en train de, Être actuellement occupé à : Être en train de dîner. Être en train de se ruiner.
 || Être en disposition, en humeur de : Je ne suis pas en train de rire.

2.2.2. Scénarios du changement linguistique

Pour expliquer l'émergence de *être en train de* dans son acception progressive, différents scénarios ont été proposés, la plupart se fondant moins sur l'attestation d'exemples tirés de corpus, que sur des théories, mécanismes, et scénarios attestés dans un cadre typologique. Pusch, notamment, part du principe que : « la construction *être en train* subit une dé-modalisation pour devenir courante, dans sa lecture progressive, vers le milieu du 19^e siècle » (2003 : 501). Il reprend ensuite le schéma de la grammaticalisation de l'aspect de Bertinetto² et s'étonne de ce que *en train de* paraît « presque totalement spécialisée dans la fonction focalisée de l'aspectualité progressive » (2003 : 506), sans pourtant être passée par une valeur durative.

Pour éclairer le cheminement diachronique de *en train de*, Mortier (2005) s'attache à retracer l'étymologie de *train* depuis le douzième siècle, pour en dériver le sens de *train* dans *en train de*. Plusieurs difficultés se présentent : d'abord, *train*, au XVII^e siècle, est extrêmement polysémique, avec huit entrées différentes au total (incluant *en train* et *tout d'un train*) dans le dictionnaire de Furetière ; ensuite, la prise en compte des autres éléments de la périphrase, *en*, *de*, *être*, et de leur contribution sémantique à l'ensemble, n'a rien d'évident. Ce travail étymologique était cependant l'invocation d'un mécanisme de grammaticalisation, déclenchée par une extension métaphorique du sens de *train* vers l'aspectualité convoquée par *en train de* :

Cette transgression peut être représentée de manière visuelle : tout comme le « *train-objet* » est une succession de wagons en mouvement, le « *train-marqueur du progressif* » exprime des moments successifs dans le temps. Au moment où *train* prend donc le sens de « mouvement » ou d'« allure », il semble se déchaîner une évolution métaphorique, que nous pouvons représenter de la façon suivante : TRAIN (marche, mouvement, vitesse, allure) > (métaphore) ÊTRE EN TRAIN DE (être au milieu du mouvement, de l'action, donc être en cours). (Mortier 2005 : 86)

Cette suggestion ne nous paraît pas très convaincante. En effet, la métaphore (d'aucuns ont préféré parler de métonymie pour qualifier cette transition ; cf. Momma 2023 : 75) manque de spécificité, et le scénario ne paraît pas rendre justice à la complexité du processus historique, laissant implicitement de côté le développement et l'influence éventuelle du sens modal. Cette mise à l'écart est explicitement confirmée dans un article ultérieur :

-
2. Ce schéma, proposé pour l'émergence du progressif dans les langues romanes, obéit à 5 étapes : (i) locatif pur (emploi statif/duratif) ; (ii) progressif I (emploi statif résiduel, emploi duratif, aspectuellement neutre) ; (iii) progressif II (emploi duratif, aspectuellement neutre) ; (iv) progressif III (emploi focalisé, strictement imperfectif) ; (v) imperfectif pur, sans valeur progressive (Bertinetto 2000). La périphrase *en train de* est associée par Bertinetto à l'étape (iv) de ce schéma, ce qui correspond bien à l'interprétation qu'en donne Gosselin (2011).

Dérivé d'un verbe de mouvement, *train* n'acquiert pas le sens d'humeur ou de disposition avant le dernier quart du XVI^e siècle, et quand bien même, il ne se rencontre pas fréquemment avant le dernier quart du XVII^e siècle. En revanche, la périphrase apparaît régulièrement avec un sens pleinement progressif à partir du XVI^e siècle. (Mortier 2008 : 10 ; notre traduction)

Ce constat l'amène à rejeter l'idée que le sens progressif actuel de la périphrase trouve son origine dans ses emplois à valeur modale.

Do-Hurinville (2007) est plus nuancé. Il retrace de façon synthétique l'évolution des formes *train*, *en train*, *mettre en train*, *être en train de*, situant l'émergence du sens progressif dans le courant du XIX^e siècle. Il suggère cependant que différents chemins de grammaticalisation ont pu concourir à ce résultat :

L'évolution de *train* vers *être en train de* semble conforme aux « chemins sémantiques », présentés par Marchello-Nizia [...] : *mouvement* > *aspect* ou *auxiliaire modal* > *auxiliaire aspectuel*, en ce sens que *train*, qui exprime un mouvement, forme *être en train de*, marquant l'aspect progressif, et que *être en train de*, au sens modal de « en disposition de », devient ensuite un auxiliaire aspectuel tel qu'on le connaît actuellement. (2007 : 33)

L'origine de la valeur modale n'est cependant pas discutée. Do-Hurinville mentionne en outre que « la périphrase *en train de* peut saisir le procès en trois phases (antérieure, initiale, médiane) » (2007 : 33), ce qui conduit à une pluralité de valeurs aspectuelles, et notamment une valeur imminentielle présente dès le XVII^e siècle. Si le tableau diachronique ainsi dressé peut paraître quelque peu embrouillé, avec divers chemins enchevêtrés et une multiplication de valeurs qui n'ont pas toutes survécu, il nous paraît cependant plus fidèle à ce qui ressort de l'examen empirique des occurrences historiquement attestées.

2.2.3. L'hypothèse parallèle

On ne saurait toutefois achever cet état de l'art sans présenter le récent travail de thèse de Momma (2023), entièrement consacré au développement diachronique de *en train de*, et qui constitue la première étude exhaustive de l'ensemble des occurrences de *en train de* issues de la base textuelle Frantext, chacune ayant été annotée suivant une grande diversité de critères (sémantiques, morphologiques, etc.). Momma distingue d'abord deux hypothèses majeures pour le développement diachronique de *en train de* : une transition directe *mouvement* > *progressif*, et une transition indirecte *mouvement* > *disposition* > *progressif*. Elle privilégie une solution que nous qualifierons de « parallèle », postulant une double transition *mouvement* > *progressif* et *mouvement* > *disposition*, les deux étant tenues pour indépendantes.

Pour battre en brèche l'hypothèse d'une source modale, Momma fait remarquer que le sens progressif est antérieur au sens modal, et toujours plus fréquent, quelle que soit la période considérée (2023 : 142). Seulement, ce constat se fonde sur son propre travail d'annotation, où chaque occurrence de *en train de* peut

prendre trois valeurs sémantiques : clairement progressive, clairement modale, ou ambiguë entre les deux – nous rapportons pour chacune, respectivement en (11), (12) et (13) un des exemples tirés de Frantext donnés en illustration dans la thèse. L’ambiguïté en (13) est attribuée au fait que l’on ignore si Gename était déjà en train de parler ou non lorsqu’on lui a intimé l’ordre de se taire : selon Momma, dans le premier cas il s’agit d’un procès en cours, dans le second d’une disposition modale. Suivant ce raisonnement, on perd donc la valeur modale dès lors que la disposition est réalisée.

- (11) Ils descendirent au rez-de-chaussée ; et elle était en train d’expliquer l’usage des cassettes, quand Sénécal, qui les avait suivis, s’interposa entre eux. (G. Flaubert, *L’Éducation sentimentale*, t. 1, 1869)
- (12) Juché sur le dossier du fauteuil, le singe privé croquait des dattes et faisait claquer ses dents ; contre les jambes du maître le chat favori se frottait en arrondissant le dos ; le nain difforme tirait la queue du singe et les moustaches du chat, dont l’un glapissait et l’autre jurait, ce qui ordinairement déridait sa majesté ; mais sa majesté n’était pas ce jour-là en train de rire. (T. Gautier, *Le roman de la momie*, 1858)
- (13) Parthénoïde tout sérieux et sage qu’il est, ne pût s’empêcher de rire de voir avec quelle ardeur Philonime sollicitait son audience ; et se promettant quelque chose de fort gay ou d’extraordinaire, obligea Gename, qui étoit assez en train de causer, de faire silence, et d’écouter Philonime. (M. de Pure, *La Prétieuse ou le Mystère des ruelles*, parties 1-2, 1656)

Outre la portion congrue allouée au sens modal, confinée à la seule virtualité, ce schème d’annotation nous semble présenter un autre écueil, à savoir que dans les contextes diachroniques, la palette de sens pris par *en train de* est possiblement plus large qu’une simple partition progressif/modal (on a mentionné notamment le sens inchoatif et le sens imminentiel). D’autre part, comme on l’a soulevé *supra*, il est fréquent de projeter sur des occurrences anciennes, au sens parfois difficile à démêler, un sens actuel duquel l’annotateur est familier. Dès lors, on peut discuter la fiabilité d’une figure qui n’est jamais que la traduction d’une certaine interprétation du sens pris individuellement par les diverses occurrences.

Quoi qu’il en soit, les résultats de l’annotation présentés dans cette thèse parlent selon nous en défaveur de l’hypothèse parallèle telle qu’elle a été conçue : en effet, dans la figure récapitulative du dénombrement des occurrences modales et progressives (Momma 2023 : 143), les courbes de fréquence des trois valeurs sémantiques (progressif, modal, ambigu) paraissent étroitement corrélées jusqu’au début du XIX^e siècle. Cela suggère que l’évolution de leur usage ne constitue pas des processus indépendants et séparés, comme le voudrait l’hypothèse parallèle, mais participent au contraire d’un même mouvement, qui s’effectue de manière transverse à la classification modalité/aspectualité.

2.3. Étude qualitative

Pour mieux comprendre le développement historique complexe de *en train de*, nous l'illustrons ici de quelques occurrences choisies et significatives, que nous tirons du corpus Frantext.

2.3.1. Premières attestations

La première attestation, datant de la toute fin du xv^e siècle, illustre la tournure *mettre en train* :

- (14) Il me desplaisoit fort de partir que je ne meisse en train de reprendre ceste paix; et veoye les maulx que vous ay dict et approucher l'yver. (P. de Commynes, *Mémoires*, t.3, 1495)

Dans cet exemple, un conflit oppose le duc de Milan et le roi de France, représenté par le duc d'Orléans ; ce dernier est basé avec ses troupes à Navarre, mais manque de provisions et la famine s'installe. Leur sort dépend de la décision du Roi, qui hésite entre une résolution pacifique qui permettrait au duc d'Orléans de lever le camp, et le maintien des troupes à Navarre pour ne pas céder devant le duc de Milan. Aucun des deux camps ne veut prendre l'initiative des pourparlers, ce qui serait avoué de faiblesse. Philipppes de Commines, jusqu'ici présent dans l'entourage du roi, est sur ces entrefaites (« estans les choses en ce train ») dépêché au château de Casal pour y régler une affaire ; c'est là qu'intervient l'occurrence. On comprend de ce contexte que « mettre en train de reprendre cette paix » signifie 'réussir à enclencher les pourparlers'. On est donc proche ici de la valeur « mettre en mouvement ».

Les exemples (15) et (16) viennent illustrer cette fois la valeur modale :

- (15) Ores advint bien tost après que icelle bonne commere allemande estoit en mey la rue au chault du soleil, fillant et se devisant avecques des aultres commeres ses voisines, et tellement qu'elle se prindrent à deviser de beaucoup de chose qui guieres ne valloient, entre lesquelles la bonne commere allemande mist les aultres en trayn de ung jour aller à Mets. (P. de Vigneulles, *Les Cent Nouvelles*, 1515)

Le contexte est suffisamment clair ici : la « bonne commère allemande » suggère à ses voisines d'aller à Metz, une idée qui emporte leur adhésion et les incline à s'y rendre. Là encore, la valeur de « mise en mouvement » reste proche, mais il s'agit d'un mouvement intérieur, d'une disposition.

- (16) Mais s'il plaint à ce bon Dieu de se servir de moy, il me relaschera et appaisera le mal qui me presse si fort que je ne puis vacquer à aulcun labeur d'importance [...]. Tant y a que, s'il ne me fait la grace de me restituer en meilleure disposition, je ne suis pas en train de jammais monter sur cheval. Pourtant encor que je feusse mandé, si ne pourrois-je bouger de la maison en tel estat. (J. Calvin, *Lettres à Monsieur et Madame de Falais*, 1543)

La signification n'est plus ici celle d'une inclination vers le fait de monter à cheval, mais d'une (in)capacité de le faire. La valeur est donc indéniablement modale, mais le sens est différent de (15).

L'exemple suivant (17) est convoqué par Mortier (2008 : 11) pour montrer que la préposition, avec un sens « pleinement progressif », est déjà établie dès le XVI^e siècle :

- (17) Jacques du Chastel, Evesque de Soissons, au voyage d'outremer que fist S.Loys, voyant le Roy et toute l'armée en train de revenir en France laissant les affaires de la religion imparfaites, print resolution de s'en aller plus tost en paradis. Et, ayant dict à Dieu à ses amis, donna seul, à la veuë d'un chacun, dans l'armée des ennemis, où il fut mis en pieces. (M. de Montaigne, *Essais*, t. 1, livres 1 et 2, 1592)

Ici, il est question de la septième croisade ; l'expédition contre l'Égypte n'a pas porté les fruits escomptés, et maintenant Saint Louis envisage de retourner en France, sans avancée majeure concernant la situation des Chrétiens en Terre Sainte – Jérusalem, en particulier, reste sous domination égyptienne. Ne pouvant tolérer cela, Jacques du Chastel, au rang des croisés, décide de se porter au-devant des troupes ennemies, et trouve la mort³. Il n'est donc pas question d'un progressif : Jacques du Chastel réalise le départ prochain de l'armée, et réagit en conséquence. La traduction contemporaine de Guy de Pernon rend d'ailleurs « voyant le Roy et toute l'armée en train de revenir en France » par « voyant que le roi s'apprêtait à revenir en France avec toute l'armée ». La valeur est ici imminente, voire modale, les deux pouvant se confondre (le départ est imminent, parce que les acteurs y sont disposés), mais elle n'est pas progressive : le procès « revenir en France » n'est pas en cours, il n'est pas même commencé.

Dans un autre exemple du même Montaigne, la valeur de *en train de* est moins évidente :

- (18) Feraulez, qui avoit [...] trouvé que l'accroist de chevance n'estoit pas accroist d'appetit au boire, manger, dormir et embrasser sa femme ; et qui d'autre part santoit poiser sur

3. L'anecdote est très vraisemblablement tirée de la *Vie de saint Louis* de Joinville : « Il avoit un moult vaillant home en l'ost, qui avoit à non mon seigneur Jaque de Castel, evesque de Soissons. Quant il vit que nos gens s'en revenoient devers Damiete, il, qui avoit grant desirrer de aler a Dieu, ne s'en vout pas revenir en la terre dont il estoit né, ainçois se hasta d'aler avec Dieu, et feri des esperons et assembla aus Turs tout seul, qui a leur espees l'occistrent et le mistrent en la compagnie Dieu ou nombre des martirs. » (Joinville 1998, p. 194, §393). De manière déterminante, à ce stade du récit, l'armée française a subi une déroute complète ; nombre de chevaliers et Saint Louis ont été faits captifs, et restitués en échange de Damiette, conquise deux ans plus tôt. Quand l'armée ainsi défaite revient de Damiette, chacun songe que la croisade s'est soldée par un échec et qu'un retour en France s'impose : c'est dans ce contexte que s'inscrit l'anecdote. Mais la décision n'est pas encore prise, et fera l'objet d'un conseil au cours duquel Joinville persuadera finalement le Roi contre l'avis de tous de rester en Terre Sainte, qu'il ne quittera que trois ans plus tard.

ses espales l'importunité de l'œconomie [...] : delibera de contenter un jeune homme pauvre, son fidele amy, abboyant apres les richesses, et luy fait present de toutes les siennes, grandes et excessives, et de celles encore qu'il estoit en train d'accumuler tous les jours par la liberalité de Cyrus son bon maistre, et par la guerre : moyennant qu'il prinst la charge de l'entretenir et nourrir honnestement comme son hoste et son amy. (M. de Montaigne, *Essais*, t. 1, livres 1 et 2, 1592)

Ici, la valeur progressive, plus particulièrement continuative, peut paraître indéniable ; elle est d'ailleurs renforcée par « tous les jours », forme spécialement courante chez Montaigne. Mais à mieux y regarder, cette occurrence coïncide mal avec les emplois actuels de la forme : Phéraulais n'est pas ici dans le procès de l'accumulation de richesses, puisque les richesses futures, « celles encore qu'il estoit en train d'accumuler », contraste avec « toutes les siennes, grandes et excessives ». La valeur sémantique exacte de *en train de*, difficile à saisir ici, est donc de l'ordre de l'à-venir, et se rapproche une fois encore de la valeur imminentielle, voire de la valeur modale (il s'agit d'une « disposition » dans laquelle Phéraulais se trouve, d'accumuler des richesses du fait de sa situation). On n'est pas très loin ici du « je ne suis pas en train de jammais monter sur cheval » de Calvin (exemple 16), qui signale une disposition non pas tant fondée sur l'humeur et l'inclination (absentes en 16) que sur une possibilité immédiatement inscrite dans la situation actuelle du locuteur.

Cette valeur modale de potentialité a priori garantie et qui ne dépend plus de l'intention ni de l'humeur se retrouve également dans l'occurrence suivante :

- (19) des hommes entierement diffamez pour leurs vices, comme sont Hercule, Socrate, Thesée, et quelques autres, sous pretexte de quelque petite action morale, qui les a rendu recommandables à la posterité, et de laquelle ils ont reçu suffisamment la récompense, d'autant qu'en faveur de telles actions ils sont venus jusques à nous, et sont en train de vivre jusques à la fin du monde. (F. Garasse, *La Doctrine curieuse des beaux-esprits de ce temps*, 1623)

À côté de ces emplois modaux, on trouve également divers emplois aspectuels, notamment avec la forme *mettre en train* (exemple 23), qui constitue l'emploi le plus fréquent de *en train de*.

- (20) Et ici vous fero y recit en quel estat se trouverent lendemain ceux du chateau, lesquels sus le point de leur partement, estoient en train d'acompaigner qui n'estoit avec eux. (C. de Taillemont, *Discours des Champs faëz. À l'honneur, et exaltation de l'Amour et des Dames*, 1553)
- (21) Le mulet, guary, se trouva assez humble pour un temps ; mais (je m'en ris encore, et j'eus ce plaisir), un samedy matin que ce vieillard, voulant aller aux champs, monta sur sa beste, qui sçavoit le chemin de sa cure, voylà qu'il est en train d'aller. (F. Béroalde De Verville, *Le Moyen de parvenir*, 1610)
- (22) Outre les trafics susdits que nous avons perdus, ou que nous sommes en train de perdre, il en reste encor deux que nous prenons de dehors. (A. de Montchrestien, *Traicté de l'œconomie politique*, 1615)

- (23) Il mit la compagnie où elle était en train de danser aux chansons et pria une de ses cousines, qui savait son dessein, d'en chanter une où chacun entre l'un après l'autre au milieu de la danse et va baiser la dame qui lui plaît le plus.
(C. Sorel, *La Sœur jalouse*, 1623)

L'exemple (20) présente une valeur proche de (17) : Laurine, une princesse de Perse doit être conduite à Antioche par son père le duc d'Orée pour y épouser le prince ; sur la route, ils font escale au château de Corvin d'Asse. Alors qu'ils doivent appareiller le lendemain, le fils de Corvin, Léontin, de connivence avec la princesse, profite de la nuit pour l'enlever et prendre la mer avec elle. L'occurrence fait immédiatement suite à cette situation : *en train d'accompagner* ne désigne alors pas un procès en cours, puisque précisément la personne que les gens du château sont censés accompagner n'est plus là.

L'exemple (21) présente une valeur inchoative, à dimension progressive. En (22), Montchrestien détaille les affaires commerciales que la France perd au profit des Flamands : ceux qu'elle a déjà perdus, ceux qu'elle perd graduellement (« C'est par ce moyen qu'ils [les Holandois]... nous soustrayent peu à peu [le trafic] de Barbarie. »), ceux qu'elle va perdre instamment (« ne faut point douter qu'en peu de temps ils ne s'attribuent, au préjudice des François, tout le negoce de ceste terre [le Canada] »). L'aspect associé à *en train de* est donc tout à la fois progressif et imminentiel. Enfin, l'exemple (23), notamment à travers le sémantisme de *mettre*, imprime à « danser aux chansons » une valeur inchoative.

Comme l'a justement indiqué Momma (2023), il apparaît donc que, avant même le plein développement du sens modal de « disposition » attaché à *en train de*, un sens aspectuel était déjà en place. Nous ne pensons pas cependant que le sens progressif contemporain était déjà établi : les usages associés à *en train de* laissent plutôt l'impression d'une matrice aspectuelle mal définie, susceptible de prendre diverses valeurs en fonction du contexte, et présentant ainsi de multiples facettes, incluant, outre une dimension progressive qui n'est pas nécessairement active, l'imminentiel et l'inchoatif.

De fait, *en train de* reste marginal : il se combine encore avec des noms, et dépend le plus souvent d'une construction, [*mettre en* {N} *de* {Vinf}], qui peut prendre d'autres valeurs (*mettre en peine de*, *mettre en liberté de*, *mettre en fait de*, *mettre en hasard de*, *mettre en devoir de*, etc.), de même que le motif [*en* {N} *de* {Vinf}] s'accommode d'une multitude de termes (*en intention de*, *en espoir de*, *en état de*, *en oubli de*, *en vouloir de*, *en volonté de*). Il est donc probable qu'avant le dix-septième et l'émergence plus claire de la valeur modale, signalée en particulier dans les dictionnaires, *en train de* n'ait pas existé comme construction enracinée en tant que telle dans les usages linguistiques, mais comme une séquence de termes rendue probable par l'existence de différents schémas, prenant dès lors son sens en contexte, faute d'être inscrite à part entière dans les inventaires cognitifs de la langue d'alors.

2.3.2. L'émergence de la valeur modale

Au XVII^e siècle, la valeur modale de *en train de* se voit largement étoffée. L'un des principaux emplois de la construction avec cette valeur se rapporte au discours : être *en train de parler*, c'est en avoir le désir, la disposition ; si donc un agent est *en train de parler*, il continuera probablement de le faire, à moins d'une interruption explicite. Les exemples (24) à (28), qui incluent des occurrences de *en train* seul dont l'usage paraît analogue, illustrent ce fonctionnement. C'est à l'aune de ces emplois qu'il faut également interpréter l'exemple (13), qui se range parfaitement dans cette série : certes Gename est bel et bien occupé à parler, mais le fait qu'il est *en train de parler* souligne sa disposition à continuer, et donc la nécessité d'y mettre un terme : il ne se taira pas tout seul.

- (24) Ce parfait amant estoit en train de dire une infinité de choses pour soutenir la gloire de Charite, mais le maistre de la maison fit cesser tous ces discours afin que l'on se mist à table, où il ne vouloit parler que de boire. (C. Sorel, *Le Berger extravagant*, 1627)
- (25) Mais le bon père était si en train, qu'on lui eût fait tort de l'arrêter, de sorte qu'il poursuivait ainsi. (B. Pascal, *Les Provinciales*, 1657)
- (26) Gelaste fut surpris de ce mot de Galathis ; mais la voyant si bien en train, il ne voulut pas l'interrompre en l'avoüant ou le deguisant. (M. de Pure, *La Prétieuse ou le Mystère des ruelles*, parties 3-4, 1658)
- (27) Je suis dans les mêmes sentimens que vous pour les définitions, monsieur. Toute la difference qu'il y a entre nous deux, c'est que je suis un peu plus occupé d'ailleurs que vous, et que vous y songez davantage que moi ; mais quand on me met en train de définir, je ne veux plus faire autre chose. (R. de Bussy-Rabutin, *Les Lettres de messire Roger de Rabutin, comte de Bussy*, t.1, 1666-1681, 1681)
- (28) Je suis bien en train de causer : il faut pourtant vous quitter. (F.-T. de Choisy, *Journal du voyage de Siam fait en 1685 et 1686*, 1687)

On notera par ailleurs que Mme de Sévigné semble affectionner particulièrement cette forme linguistique (exemples 29 à 31), qu'elle emploie également dans des tournures négatives (exemples 32 à 34). Dans l'exemple (30), la mention de « [ma plume] est bien en train de trotter » est un artifice qui lui sert de transition : puisqu'elle est disposée à écrire, elle poursuit donc sa lettre, cette fois sur un tout autre sujet. Il convient de souligner ici l'analogie fonctionnelle entre *en train de* et *en humeur de*, lesquels se trouvent en conjonction dans les exemples (31) et (34), ce qui montre bien la proximité sémantique entre les deux. On rapprochera en outre de cette série les exemples (8) à (10), signe que cet emploi modal reste vivace jusqu'à la première moitié du XIX^e siècle au moins, en particulier vis-à-vis de l'inspiration à écrire ou produire un contenu linguistique.

- (29) Il était bien en train de discourir aujourd'hui. Mme De La Troche et le prieur de Livry étaient ici ; il s'est bien diverti à leur prouver tous les attributs de la divinité. (Mme (de) Sévigné, *Correspondance*, t. 1, 1646-1675, 1675)
- (30) J'en ai vu encore plusieurs autres d'elle [il s'agit de lettres, écrit par la mère Angélique du Port-Royal], et bien plus belles, et bien plus justes ; ceci est un billet écrit à course de plume. La mienne est bien en train de trotter. (Mme (de) Sévigné, *Correspondance*, t. 2, 1675-1680, 1680)
- (31) Je ne sais ce qui me tient que je ne vous conte le procès dont il est question, tant je me sens en train de discourir, mais je m'arrête, car il se pourrait fort bien faire que vous ne seriez pas en humeur de m'écouter, et je vous veux plaire. (Mme (de) Sévigné, *Correspondance*, t. 2, 1675-1680, 1680)
- (32) Adieu, ma très chère ; je ne suis pas bien en train de vous parler d'autre chose. (Mme (de) Sévigné, *Correspondance*, t. 2, 1675-1680, 1680)
- (33) Vous voulez que j'écrive à M De Vivonne : eh, bon dieu ! N'est-il pas trop bien payé de vous avoir vue, de vous avoir régaler ? Ce serait donc pour se réjouir avec lui de ce qu'il est plus raisonnable cette année que l'autre qu'il faudrait lui faire un compliment. J'en avais tantôt commencé un. Ma plume n'était pas en train ; j'ai tout planté là. (Mme (de) Sévigné, *Correspondance*, t. 2, 1675-1680, 1680)
- (34) Ainsi nous ne sommes point en train ni en humeur de faire des promenades extravagantes. (Mme (de) Sévigné, *Correspondance*, t. 3, 1680-1696, 1696)

L'usage de *en train de* pour communiquer une disposition interne ou une inclination subjective ne se résume pas au seul contexte de la verbalisation, comme en attestent les exemples (35) à (37). Dans l'exemple (35), *trancher de l'homme d'importance* signifie vraisemblablement affecter d'être un homme d'importance (sens attesté dans le dictionnaire de l'Académie de 1694) : Mercure (le locuteur de l'occurrence) reproche donc au personnage de Sosie (auquel il s'adresse) d'adopter la posture d'un important – d'où sa réaction de vouloir le souffleter. Le sens de *en train de* relève ici de l'inclination, de la disposition, mais cette inclination est prêtée à Sosie par Mercure ; dès lors, c'est l'activité qui souligne la disposition modale, laquelle se trouve ainsi reléguée au second plan de la représentation mise en place par l'énoncé. Nous retrouverons ce genre d'occurrence dans la section suivante.

- (35) Tu montres de l'esprit, et je te vois en train De trancher avec moi de l'homme d'importance. Il me prend un desir, pour faire connoissance, De te donner un soufflet de ma main. (Molière, *Amphitryon*, 1668)
- (36) On prend cette vipère, on lui coupe la tête, la queue, on l'ouvre, on l'écorche, et toujours elle remue. Une heure, deux heures, on la voit toujours remuer. Nous comparâmes cette quantité d'esprits, si difficiles à apaiser, à de vieilles passions, et surtout celles de ce quartier. [...] On croit que quand on leur arrache le cœur, c'en est fait, qu'on n'en entendra plus

parler. Point du tout. Elles sont encore en vie, elles reviennent encore. Je ne sais pas si cette sottise vous paraîtra comme à nous, mais nous étions en train de la trouver plaisante ; on en peut faire souvent l'application. (Mme (de) Sévigné, *Correspondance*, t. 2, 1675-1680, 1680)

- (37) Je m'en vais, pendant que je suis en train, vous expédier la Cochinchine. (F.-T. de Choisy, *Journal du voyage de Siam fait en 1685 et 1686*, 1687)

Comme dans l'exemple (16), cette disposition interne peut par ailleurs ne pas être subjective, mais se rapporter aux caractéristiques générales du sujet, ainsi qu'en atteste l'exemple (38). On notera le jeu qui s'opère entre le caractère physiologique de la sudation, sur laquelle Mme de Sévigné n'a évidemment aucun contrôle, et une forme d'intention néanmoins prêtée, « elle dit que j'aime à suer » ; *en train de* participe vraisemblablement du même jeu.

- (38) Je suis tellement en train de suer que je sue toujours, et la bonne D'Escars n'ose me proposer d'ôter des habits, parce qu'elle dit que j'aime à suer. (Mme (de) Sévigné, *Correspondance*, t. 2, 1675-1680, 1680)

Cette large attestation du sens modal au XVII^e siècle n'exclut pas pour autant la persistance des emplois aspectuels et notamment progressifs déjà entrevus précédemment, quoiqu'ils soient de loin moins fréquents. On retrouve notamment un sens inchoatif – *être en train de*, c'est être en bonne voie de, notamment dans le sens d'une convalescence. C'est ce qu'illustrent les exemples (39) à (41) :

- (39) Mais vous estes en grand train de m'obliger de toute façon, et moy en estat de mourir ingrat, puisque je n'ay pas moyen de me revancher de tant de courtoisies et de bienfaits que j'ay reçeus de vous depuis tantost six ans (G. Patin, *Lettres*, t. 1, 1630-1649, 1649)
- (40) Je traite fort sérieusement la santé du chevalier, et je verrai les commencements de ses remèdes, et le laisserai en bon train avant que de partir. (Mme (de) Sévigné, *Correspondance*, t. 2, 1675-1680, 1680)
- (41) Je vous ai souhaité un lot à la loterie, pour commencer à rompre la glace de votre malheur. [...] Cela donc vous aurait remis en train d'être moins malheureux, mais je crois que ma nièce de Sainte-Mariele saurait, et qu'elle me l'aurait dit. (Mme (de) Sévigné, *Correspondance*, t. 3, 1680-1696, 1696)

Enfin, l'exemple (42), là encore tiré du riche éventail d'emplois que fait Mme de Sévigné de la construction, peut paraître investi d'une pleine valeur progressive :

- (42) Je comprends, ma chère enfant, cet ouvrage de deux mois que vous avez à faire cet hiver à Aix. Il paraît grand et difficile à le regarder tout d'une vue, mais quand vous serez en train d'aller et de travailler, étant tous les jours si accablée de devoirs et d'écritures, vous trouverez que, malgré l'ennui et la fatigue, les jours ne laissent pas de passer bien vite. (Mme (de) Sévigné, *Correspondance*, t. 3, 1680-1696, 1696)

Cependant, dans ce dernier exemple, la valeur aspectuelle inchoative explique tout aussi bien l'occurrence : « quand vous serez en train d'aller et de travailler »

peut aussi bien s'interpréter par « une fois que vous vous y serez mise » : une fois que vous aurez commencé l'ouvrage, vous serez absorbée par le travail et le temps passera vite. Les deux lectures sont possibles et il est difficile de voir dans *en train de* son emploi actuel qui permet de présenter comme tel un procès dans sa durée.

Ainsi, au xvii^e siècle, l'usage principal de la construction est un usage à valeur modale, signalant la disposition ou l'inclination, la plupart du temps subjective, pour se prêter à une certaine action. Cette valeur, associée à la majorité des occurrences de Frantext pour cette période, correspond tout à fait aux définitions proposées dans les dictionnaires de l'époque, signe d'une prise de conscience métalinguistique par les lexicographes de l'unité de cette séquence. Quant à la valeur aspectuelle, elle subsiste, mais reste mal définie ; et si un sens progressif peut parfois se faire jour, celui-ci est plus imputable aux variations contextuelles d'un sens encore flou, qu'à une valeur sémantique pleinement intégrée à l'usage de cette construction.

2.3.3. Le tournant du xviii^e siècle

Le xviii^e siècle semble constituer une période charnière dans l'évolution de la forme ; c'est en effet dans l'édition 1798 du dictionnaire de l'Académie qu'apparaît pour la première fois une valeur progressive, puisque la définition précise qu'*être en train de jouer, de courir*, c'est aussi jouer, courir actuellement ; dans cette veine Francis Wey, en 1845, glose *en train de* par « être occupé à » (cité par Gougenheim 1929 : 64). Gougenheim identifie en tout cas la période s'étendant entre la fin du xviii^e et le début du xix^e comme une période de transition, notant que l'interprétation exacte de la forme est souvent difficile. Nous présentons ici quelques occurrences pour illustrer cette versatilité sémantique.

La première occurrence (43) semble fonctionner suivant l'usage progressif actuel : *en train de rire* introduit une activité qui sert de cadratif, de toile de fond à un événement qui vient l'interrompre. Cependant, la seconde occurrence (44), également avec *rire*, de la même année et du même auteur, vient nuancer cette lecture : elle exprime certes un procès en cours, là encore interrompu de manière inopinée ; mais en même temps elle ne contredit pas une lecture modale, puisque « nous étions le plus en train de rire » correspond étroitement à « notre joie », signe d'une disposition, d'un état d'esprit des participants. Ces occurrences au fonctionnement double, dont l'emploi dans la séquence narrative font ressortir le jeu entre la disposition à faire et l'activité en cours, nous paraissent constituer des *bridging contexts* tels que les définit Heine (2002) dans sa conception du processus de grammaticalisation. Une troisième occurrence (45) illustre par ailleurs, chez cet auteur toujours, une valeur modale toujours vive.

- (43) La conversation ne tarda guère à devenir galante et badine ; elle s'échauffa, et l'arrivée de quelques dames des plus éveillées du couvent ne la refroidit point. [...] Nous étions tous bien en train de rire quand on vint dire à madame l'abbesse que Don Antonio De Miras allait paraître au parloir [...]. (A.-R. Lesage, *Histoire de Guzman d'Alfarache*, t. 4, 1732)
- (44) Le repas fut assaisonné de mille discours agréables. Fabrice surtout, qui savait donner de l'enjouement à la conversation, divertit fort la compagnie. [...] Dans le temps que nous étions le plus en train de rire, notre joie fut tout à coup troublée par un événement imprévu. Il entra dans la chambre où nous soupions un homme assez bien fait, suivi de deux autres de très mauvaise mine. (A.-R. Lesage, *Histoire de Gil Blas de Santillane*, Livres 1-12, 1732)
- (45) Je l'avoue à ma honte, ils me firent connoître que je n'avois pas, à beaucoup près, fait un livre qu'on ne pouvoit critiquer. J'en eus tant de dépit, que je jetai mon histoire de la poésie au feu ; et tandis que j'étois en train de brûler, j'abandonnai aux flammes tous mes papiers, en faisant serment de ne plus écrire. (A.-R. Lesage, *La Valise trouvée*, 1740)

L'occurrence (46), due à Diderot, est également tout à fait compatible avec une lecture progressive : *en train de* pose un procès que le passage de la « beauté » vient interrompre. Le sens modal est d'ailleurs d'autant plus effacé que l'action introduite (les génuflexions) correspond à un cérémoniel précis (prescrites par la loi), et n'engage pas a priori la disposition subjective du protagoniste, pas plus que ses aptitudes générales ou une propension d'aucune sorte. En somme, il s'agit là d'un *switching context*, pour reprendre la terminologie de Heine (2002) : le sens progressif constitue bien l'interprétation principale de l'énoncé, et le sens modale n'est que difficilement accessible ici.

- (46) Un jour qu'il assistait dans le temple au culte public de la grande pagode, et que, selon le cérémonial usité, il était en train de lui faire les dix-sept génuflexions prescrites par la loi, la beauté dont il était épris vint à passer, et lui lança un coup d'œil accompagné d'un souris, qui le jetèrent dans une telle distraction, qu'il perdit l'équilibre, donna du nez en terre, scandalisa tous les assistants par sa chute, oublia le nombre des génuflexions et n'en fit que seize. (D. Diderot, *Les Bijoux indiscrets*, 1748)

Le cas de Diderot est par ailleurs particulièrement intéressant. Sa correspondance avec Sophie Volland offre ainsi 13 occurrences de *en train de*, qui ont l'avantage d'être toutes du même auteur et de s'inscrire toutes dans le même registre linguistique. Ces occurrences (rangées ici chronologiquement) dessinent un panorama assez exhaustif de la diversité des emplois de *être en train de* et *mettre en train de* : un usage modal plein (exemple 49, qui signale bien la disposition subjective), un usage modal conventionnel à tendance formulaïque, qui relève plus du code d'écriture que de l'expression véritable d'une propension particulière à écrire (exemples 48, 53, 54), un usage modal plus ambigu enfin où le sens progressif est réhaussé par l'articulation action en cours/interruption (exemples 47 et 50, à rapprocher des exemples 43 et 44). On trouve également l'emploi « en bonne voie » dans un contexte de guérison (exemple 52 ; cf. exemple 40) et des emplois

progressifs clairs (exemple 51), similaires à l'exemple (46), qui viennent prolonger l'extension sémantique des exemples 47 et 50 en effaçant cette fois toute dimension subjective, rendant particulièrement difficile une interprétation modale.

- (47) Il avoit été de la promenade ; il devoit être du souper ; mais il ne vint qu'après. Nous avions dévoré, les femmes surtout ; nous étions en train de dire des folies et d'en faire lorsque le cher petit prêtre arriva. (D. Diderot, *Lettres à Sophie Volland*, t. 1, 1759-23 sept. 1762, 1762 [1759])
- (48) Mais puisque je suis en train de vous écrire toutes nos minuties, il ne faut pas que j'oublie de vous raconter comme quoi Pouf, le fils de Thisbé, qui avoit fait concevoir de lui de si grandes espérances, a jetté la division parmi nous. (D. Diderot, *Lettres à Sophie Volland*, t. 1, 1759-23 sept. 1762, 1762) [1760])
- (49) Le courrier prochain m'apprendra peut-être votre départ. Ne vous attendre que pour les derniers jours du mois, je ne scaurois. Vous m'avez mis en train d'espérer. S'il nous est permis d'aller au-devant de vous, vous nous le direz apparemment. (D. Diderot, *Lettres à Sophie Volland*, t. 1, 1759-23 sept. 1762, 1762 [1760])
- (50) Hier j'eus une prise très-forte avec le maître de la maison. On étoit en train de déchirer un honnête homme de notre connoissance : c'est Crammer, libraire, de Genève. J'interrompis finement la médisance, et je dis que je souffrois avec impatience qu'on parlât mal d'un honnête commerçant étranger, par la mauvaise opinion que cela pouvoit me donner de tout honnête commerçant françois. (D. Diderot, *Lettres à Sophie Volland*, t. 1, 1759-23 sept. 1762, 1762 [1761])
- (51) Je lui ai fait lire votre rêve, à cette petite sœur, et elle trouve que vous rêvez avec plus de sens commun que les autres n'en ont éveillés ; et puis nous étions en train de discuter l'affaire des maisons, lorsque le Perronet arriva. Je crus qu'il étoit honnête de laisser ensemble des gens qui ne s'étoient vus depuis si longtems, et qui devoient avoir beaucoup de choses à se dire, toutes celles qu'ils s'étoient écrites. (D. Diderot, *Lettres à Sophie Volland*, t. 2, 26 sept. 1762-1774 [1765])
- (52) D'Amilaville est arrivé le col un peu gros encore, mais en train de guérir ; pourvu que la vie de Paris ne s'y oppose, ni femmes, ni veilles, ni table, ni vin ! (D. Diderot, *Lettres à Sophie Volland*, t. 2, 26 sept. 1762-1774 [1765])
- (53) Puisque je suis en train de vous rendre compte de mon tems, il ne faut pas oublier de vous dire que j'ai été une fois à Mousseau, où la journée se seroit assez agréablement passée, si le petit ouragan Naigeonne s'étoit brouillé avec deux de ses amis à propos d'une question de musique ! (D. Diderot, *Lettres à Sophie Volland*, t. 2, 26 sept. 1762-1774 [1768])
- (54) C'est un cousin de M De Malesherbes, et par conséquent un peu parent de l'abbé. Puisque je suis en train de vous vuidier mon sac, je vous dirois que M De Falbaire, l'auteur de l'honnête criminel, fort lié avec un musicien médecin appelé Laborde, un des tenants de la Guimare, se trouva, je ne scais comment, engagé dans un dîner chez cette belle demoiselle-là. (D. Diderot, *Lettres à Sophie Volland*, t. 2, 26 sept. 1762-1774 [1769])

On peut donc dire que Diderot résume à lui seul toute la richesse fonctionnelle et sémantique de *en train de* au XVIII^e, et qu'il l'exploite à son plein potentiel. C'est par ailleurs révélateur : l'usage qu'il en fait est à la fois extensif et pionnier : on ne retrouve guère d'autre occurrence au XVIII^e siècle exprimant avec la même clarté l'emploi progressif ultérieur. Il se pourrait donc que c'est précisément parce que Diderot étend l'emploi de cette construction au maximum, qu'il le prolonge dans un domaine auquel elle n'est pas encore associée de façon conventionnelle.

Diderot mis à part, l'usage au XVIII^e siècle paraît néanmoins dominé par la valeur modale. L'exemple (55), de Voltaire, illustre encore une fois la disposition, mais avec un sens de résolution plus qu'avec un sens d'humeur ou d'occupation – la lecture progressive est en tout cas exclue. On pourrait presque lire, dans cette occurrence, l'idée d'une « permission » (vu tout ce qu'il m'a fait, je suis bien dans une situation où je peux le tuer sans être moralement en tort), ou d'un « devoir » (il faut que je le tue). La valeur modale est également soulignée en (56), qui offre une variante de l'emploi formulaïque annonçant la poursuite d'une lettre (et généralement, un glissement thématique, comme c'est également le cas ici). La construction est, de manière inhabituelle, employée avec un nom, *sincérité*, dont le sémantisme met bien en valeur la disposition du sujet.

- (55) Voici dans ce moment ce qui se passa dans l'âme de Candide, et comment il raisonna : « si ce saint homme appelle du secours, il me fera infailliblement bruler ; il pourra en faire autant de Cunégonde ; il m'a fait fouetter impitoyablement ; il est mon rival ; je suis en train de tuer, il n'y a pas à balancer. » Ce raisonnement fut net et rapide, et sans donner le tems à l'inquisiteur de revenir de sa surprise, il le perce d'outre en outre, et le jette à côté du juif. (Voltaire, *Candide, ou l'Optimisme*, 1759)
- (56) La privation des grâces est un défaut que les femmes ne pardonnent point, même au mérite, et j'ai peur que Julie n'ait été femme une fois en sa vie. Puisque je suis en train de sincérité, je te dirai encore, ma jolie prêcheuse, qu'il est inutile de vouloir donner le change à mes droits, et qu'un amour affamé ne se nourrit point de sermons. (J.-J. Rousseau, *La Nouvelle Héloïse*, 1761)

L'exemple (57) illustre enfin la nuance sémantique que l'on peut facilement gloser par « en voie de » ; on notera qu'ici *train* est en outre qualifié par un adjectif, de même que *voie* peut l'être. Cet emploi, notamment rencontré en (40), (41) et (52), se rattache également à l'exemple (14), la plus ancienne occurrence de *en train de* attestée dans Frantext. S'il s'inscrit bien dans l'idée d'une disposition (des choses, non du sujet, arrangées pour mener à un résultat désormais attendu), il est probablement antérieur à l'emploi strictement modal, et contribue à en éclairer les contours.

- (57) Mais dès le même soir, M. Pombelins alla voir maître Molé pour lui apprendre que leurs projets étaient en bon train de réussir. (N. Rétif de la Bretonne, *La Vie de mon père*, 1778)

4. On doit par ailleurs à Diderot 22 des 158 occurrences de *en train* au XVIII^e siècle, soit près de 14% du total, alors que sa contribution au corpus de ce siècle n'est que de 4%.

Au tournant du XIX^e siècle, la valeur progressive se fait plus fréquente, sans que la valeur modale disparaisse dans l'immédiat. Les deux ne s'excluent pas, même pour un individu donné dans des contextes d'emploi très similaires, et c'est essentiellement la situation d'énonciation et le travail d'interprétation qui permettent de distinguer l'emploi modal (58) de l'emploi progressif (59), signe de la collusion entre les deux sens à cette époque, caractéristique des processus de grammaticalisation (cf. le phénomène de *layering* ; Norde & Beijering 2014 : 410).

- (58) Je te quitte, ma bonne amie, pour aller faire la lettre à M. Morel des Jardins ; je ne suis guère en train d'écrire, et cette lettre sera sans doute bien mal faite. (A.-M. Ampère, *Correspondance*, année 1802, 1802)
- (59) Il m'a été impossible de t'écrire ce matin par Pochon. Que je pestais contre moi-même de n'avoir pas préparé la lettre hier au soir ! Mais j'étais en train de recopier et de corriger le mémoire que je dois porter à Lyon à M. Delambre. (A.-M. Ampère, *Correspondance*, année 1803, 1803)

L'usage de *en train de* au XVIII^e siècle est donc caractérisé par une grande variabilité sémantique, avec des emplois qui peuvent, les exemples de Diderot ou Ampère l'ont montré, coexister, voire se confondre, dans l'idiolecte d'un individu donné. Il en ressort l'existence d'un continuum sémantique entre la disposition, l'inclination d'une part, et l'occupation, le procès duratif d'autre part. Il y a comme télescopage entre ce que le sujet est porté à faire et ce qu'il fait effectivement, et les ambiguïtés autant que la variabilité sémantique qui en découlent paraissent bien caractéristiques des contextes de transition identifiés par Heine (2002) lors des processus de grammaticalisation.

2.3.4. Synthèse du développement de la valeur progressive de *en train de*

On peut esquisser l'évolution sémantique de *en train de* comme suit. Au XVI^e siècle, la valeur dominante de *en train de* est probablement aspectuelle et plus particulièrement imminentielle ; *être en train de X*, c'est être pris dans un processus, un mouvement, un courant des choses, qui va mener à X. À cette valeur imminentielle s'ajoute également une autre valeur, inchoative voire progressive, signe que la fonction aspectuelle n'est pas encore entièrement dégagee : *être en train de perdre un trafic* (exemple 22), cela peut signifier tout à la fois qu'on l'a déjà en partie perdu (et que cela se poursuit), ou bien que cela pourrait facilement advenir. En somme, *en train de* pointe vers la direction que prend le cours des choses, direction caractérisée autant par ce vers quoi l'on tend (d'où une valeur possible de futur ; voir exemples 18 et 19), que par la tournure déjà prise par l'enchaînement des événements.

De cette valeur imminentielle, on passe à une valeur modale : si X est bientôt en cours mais ne l'est pas encore, c'est, d'un point de vue plus résolument subjectif, parce que l'agent est dans la disposition de X, qu'il est porté vers X. De même *mettre*

en train quelqu'un (Y) *de faire quelque chose* (X), c'est selon le cas agir pour que Y fasse X (valeur inchoative, exemple 23), ou bien induire chez Y une inclination vers X (valeur modale, exemple 15). Cette disposition peut alors s'étendre non pas seulement aux prémices du procès, mais à toute son étendue. Il y a en somme concomitance entre la propension à faire quelque chose, et continuer à le faire : je suis en train d'écrire, donc je continue ma lettre un peu plus. Il y a donc déplacement du sémantisme de *en train de* vers l'intérieur du procès plutôt que vers son amorce.

Au XVIII^e siècle, une première dé-subjectivation s'opère : *être en train de X* n'implique plus nécessairement un engagement subjectif de l'agent, simplement sa participation à X ; c'est le cas lorsque Diderot « discute d'une affaire » avec sa sœur, dans le courant d'une conversation (exemple 51). Le procès introduit par *en train de* peut alors simplement servir de cadre à un événement adventice. Cette dé-subjectification de l'emploi de *en train de* se traduit quantitativement par une hausse du nombre de sujets inanimés (on passe de 4% au XVIII^e siècle à 11% au XIX^e puis 17% au XX^e ; cf. Momma : 108), ainsi que par une hausse de la proportion d'emplois à la 3^e personne du singulier par rapport à la 1^{re} personne du singulier (1.1 au XVIII^e, 2.8 au XIX^e, 3.1 au XX^e ; cf. Momma : 98). Si la subjectivité s'efface, l'idée de « procès en cours », et donc l'aspect progressif, ressort.

Cependant, cette trame est un canevas dont les emplois de *en train de* ne cessent de déborder. Comme nous l'avons suggéré dans la section 1, ces sens qui nous servent de guide pour retracer le parcours historique de la forme constituent des prototypes ; mais le domaine sémantique recouvert par les usages de *en train de* accommode une diversité fonctionnelle dont il est difficile de mesurer l'exacte étendue. Les occurrences individuelles permettent de suggérer cette dispersion sémantique, malgré le caractère quantitativement limité des données disponibles. Elles témoignent d'une forme à l'emploi flexible, mais dont l'occupation sur le territoire sémantique tend à se déplacer, et c'est ce mouvement plus général que nous avons tenté de retracer ici à grands traits.

3. Étude quantitative sur corpus

Nous abordons à présent notre étude sur corpus, basé sur plus de 12 000 occurrences de *en train de* tirées de la base textuelle Frantext⁵. L'objectif de cette étude est de

5. La requête CQL utilisée est la suivante [lemma="en"%c] []{0,1} [lemma="train"|word="trayn"] []{0,1} [lemma="de"] [pos!="VINF"]{0,1} [pos!="VINF"]{0,1} [pos="VINF"]. Nous nous restreignons ainsi aux emplois suivis d'un verbe à l'infinitif, et admettons les occurrences où *train* est modifié par un adjectif, ainsi que les occurrences où ce vocable est immédiatement suivi d'un adverbe. Les deux positions libres avant le verbe ménagent également la possibilité d'y intercaler un clitique objet ou un pronom réflexif.

mettre à jour des dynamiques de fréquence qui permettraient d'éclairer l'évolution sémantique de la construction.

3.1. Courbes en S

Pour appréhender quantitativement l'usage d'une forme en diachronie, différentes mesures de fréquence sont disponibles, quoique étroitement liées entre elles (Feltgen 2022, 2024) : fréquence d'occurrence (le nombre total d'occurrences rapporté à la taille du corpus, éventuellement « par million de mots »), fréquence de types (le nombre de verbes différents rencontrés dans le schéma libre de la construction) et prévalence (la proportion de textes dans lesquels la forme est attestée⁶). Ici, nous comptabilisons les occurrences par décennie, et toutes les quantités concernées font l'objet d'une procédure de lissage, à savoir une moyenne glissante sur cinq points de données. Sont présentées Figure 1 les courbes de fréquence d'occurrence et de prévalence pour toute la période d'attestation de *en train de*.

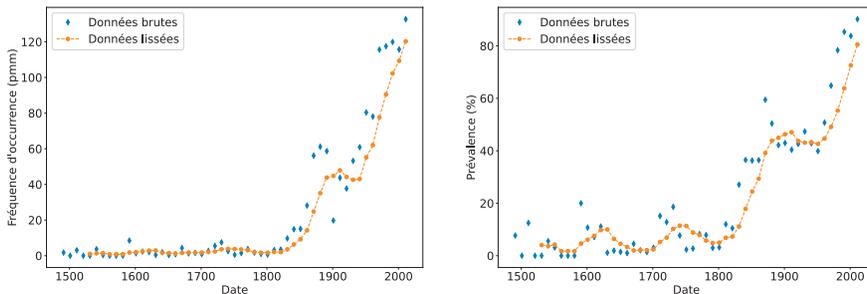


Figure 1. Fréquence d'occurrence (à gauche) et prévalence (à droite) de *en train de*

Notre première occurrence date de 1495 – voir *supra* l'exemple 14. Jusqu'à la fin du XVII^e siècle, l'emploi de *en train de* est trop sporadique pour que les variations de fréquence prennent le pas de façon fiable sur les variations individuelles ; ainsi, si la forme observée est particulièrement enracinée dans l'idiolecte d'un auteur représenté (comme c'est le cas pour Mme de Sévigné ou Diderot), cela conduit à une augmentation de fréquence qui reflète cette appétence individuelle plus qu'une tendance générale dans l'usage de la forme. Autrement dit, la fréquence d'occurrence telle que mesurée sur le corpus reste trop largement tributaire de

6. Il serait certainement plus approprié de considérer la proportion des auteurs pour lesquels la forme est attestée ; néanmoins, le caractère anonyme de certains textes, et les difficultés de traitement liées, dans Frantext, à l'encodage différencié du nom des auteurs dans le corpus et les occurrences, en rendent la mise en œuvre difficile.

la représentation ou non de certains auteurs. De même, le XVIII^e siècle présente encore de fortes variations idiolectiques ; pour la décennie 1731-1740 notamment, Alain-René Lesage produit 22 des 39 occurrences rencontrées, ce qui provoque un pic localisé de fréquence d'occurrence.

La prévalence, en revanche, est mieux appropriée pour minimiser les variations individuelles, puisque même si un auteur se révèle très friand d'une forme et l'utilise avec une propension qui le distingue de l'usage général, chaque texte ne comptera qu'une seule fois dans le calcul de la prévalence. Cette quantité connaît ici une hausse et une baisse successives (et la fréquence d'occurrence suit un motif très comparable) ; nous avons cherché à les appréhender toutes deux par un même modèle, à savoir deux courbes en S successives. Le résultat de ce modèle est donné Figure 2. À noter que la courbe en S peut être croissante (partir d'une basse fréquence pour arriver à une haute fréquence) ou décroissante (l'inverse), et que ce caractère n'est pas spécifié a priori dans le modèle, mais résulte de la calibration automatisée de ses paramètres permettant de rendre compte des données.

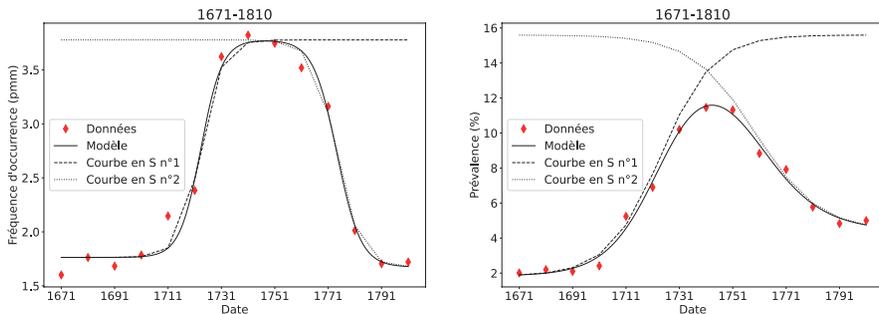


Figure 2. Modélisation par une double courbe en S de la fréquence d'occurrence (à gauche) et de la prévalence (à droite) sur la période 1671-1810.

Ce motif est particulièrement inhabituel. Si la courbe en S, plutôt que de se stabiliser sur un plateau, est régulièrement suivie par un léger décrochage de fréquence (Feltgen *et al.* 2017), ce renversement complet, suivant une courbe en S décroissante, est tout à fait rare. Si l'on accepte que ces variations d'usage reflètent des mutations sémantiques, et si l'on veut bien se risquer à interpréter un motif fragile du fait des faibles fréquences mises en jeu, l'ensemble du motif pourrait signaler tout à la fois qu'un développement du sens modal a bien lieu dans la première moitié du XVIII^e, mais que l'usage recule presque immédiatement après. On ne peut imputer cela aux concurrents directs de *en train de* : les constructions [*envie de Vinf*], [*en humeur de Vinf*] et [*enclin à Vinf*] partagent toutes une même tendance à décliner dans la deuxième moitié du XVIII^e siècle. Ce phénomène mériterait d'être éclairci.

À partir du XIX^e siècle, *en train de* connaît une fortune considérable ; dans la décennie la plus récente (2011-2020), la forme se rencontre dans 90% des textes du corpus. Cependant, il apparaît de manière frappante que cette hausse s’effectue en deux étapes marquées. Il est là encore possible de saisir l’ensemble à l’aide de deux courbes en S successives, modèle que nous rapportons Figure 3 pour la fréquence d’occurrence. Le même motif en double courbe en S se retrouve également Figure 4 pour la prévalence et la fréquence des types, que nous calculons ici en calculant le nombre de types moyen par texte présentant au moins une occurrence de la forme pour chaque décennie⁷.

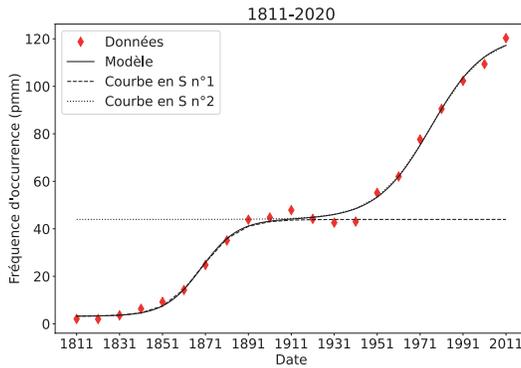


Figure 3. Modélisation par une double courbe en S de la fréquence d’occurrence sur la période 1811-2020.

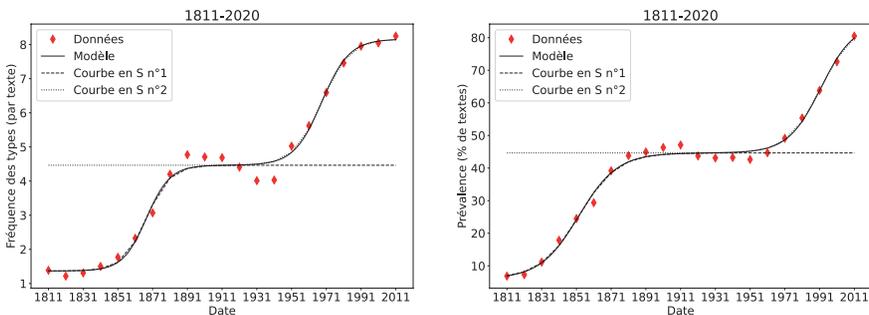


Figure 4. Modélisation par une double courbe en S de la fréquence des types (à gauche) et de la prévalence (à droite) sur la période 1811-2020.

7. Il existe d’autres manières de construire une fréquence des types tenant compte des variations de taille du corpus : par exemple en calculant la fréquence des types sur des échantillons aléatoires de corpus de taille fixe (Perek 2018) ou en exploitant la relation de Herdan entre types et occurrences (Feltgen 2024). Mais toutes ont en commun avec la nôtre que la fréquence des types résultante demeure fortement déterminée par la fréquence d’occurrence.

Arrivés à ce stade de l'analyse, nous sommes en mesure de discuter de divers éléments concernant la polysémie de *en train de*. Tout d'abord, le sens modal, qui s'ancre au XVIII^e, semble décliner aussitôt après, pour des raisons qui restent obscures. De fait, ce sens n'est plus présent dans l'usage actuel, même à titre marginal. Le sens aspectuel, progressif, se développe donc bien à partir du XIX^e seulement, même si l'extension du sens modal au XVIII^e le préparait déjà. Cependant, ce développement ne se fait pas de manière uniforme, mais en deux étapes marquées. Il est donc possible que le sens aspectuel progressif recouvre deux nuances, qui se seraient développées successivement en diachronie, sans que la seconde ne vienne concurrencer la première – elle s'y serait simplement ajoutée, comme on le voit bien sur les courbes de fréquence. Pour tester cette hypothèse, nous allons maintenant essayer de comprendre ce qui se joue dans ces deux étapes, et ce qui les distingue.

En d'autres termes, nous faisons l'hypothèse que la préposition *en train de*, en français contemporain, est bel et bien polysémique ; qu'elle présente deux sens majeurs, caractérisés par des mouvements diachroniques successifs et séparés s'inscrivant au cours des deux derniers siècles ; que ces deux sens présentent une valeur aspectuelle ; que le premier correspond au progressif, bien identifié ; que le second s'est développé à partir de lui, et reste à préciser.

3.2. Colexèmes

Dans tout ce qui suit, nous nous focalisons exclusivement sur la période 1811-2020 (après moyenne glissante, ce qui signifie que les données à partir de 1771 rentrent en ligne de compte), et considérons que la première phase court jusqu'en 1910. L'usage de *en train de* reste dominé par le type *faire* sur toute la période, que ce soit pendant la première phase (13% des occurrences) ou la seconde (9% des occurrences), aucun autre type n'arrivant près de ces chiffres. En outre, d'après le travail d'annotation de Momma (2023 : 129), la proportion entre les différents *Aktionsarten* n'évolue pas entre le XIX^e et le XX^e siècle, ce qui permet d'exclure cette piste pour expliquer une éventuelle différence sémantique.

La fréquence des types connaissant également une double augmentation « en S », il peut être tentant de considérer les types apparaissant lors de la seconde phase pour inférer la nuance sémantique correspondante. En se restreignant aux types apparaissant plus de 10 fois dans le corpus (un choix qui prévaudra dans toute la suite de cette section), on obtient la liste suivante : *cuire, lever, balayer, dormir, naître, remonter, rédiger, transformer, crier, dessiner, gravir, oublier, bouffer, mentir, sucer, surveiller, trahir, enfiler, téléphoner*. Notons que nous ne distinguons pas ici les formes régulières et les formes réflexives (e.g. *transformer* inclut *se transformer*). Cette liste est cependant problématique ; d'une part, elle inclut des types qui font surface du fait des développements techniques (*téléphoner*) ou peuvent s'expliquer par des changements de registre

rencontrés en littérature (*bouffer*). D'autre part, le premier sens de *en train de* peut très bien rester productif lors de cette seconde phase et sanctionner de nouvelles associations colexicales ; en l'état, il est donc impossible de déterminer lesquels de ces termes se rattachent à quelle nuance. Autrement dit, si un nouveau type est recruté lors de la seconde phase, on ne peut pas savoir d'emblée s'il est associé au sens déjà construit à l'issue de la première phase, ou s'il est associé au sens dont l'émergence caractérise la seconde phase, le premier sens restant présent et actif.

Une seconde possibilité est la suivante. Chaque association colexicale (ex. *en train d'arriver*) suit sa propre trajectoire, en termes de fréquence d'occurrence. On peut dès lors calculer la similarité entre cette trajectoire individuelle et la trajectoire de la construction, pour chacune des deux phases (ici à l'aide du coefficient de corrélation r de Pearson). Notons que, pour assurer l'indépendance des deux mesures (d'autant plus important dans le cas de collocatifs très fréquents comme *faire*), on peut considérer la trajectoire de la construction en termes de fréquence des types, ce que nous avons fait, même si cela n'affecte pas les résultats. Parmi les 203 types apparaissant plus de 10 fois dans le corpus, 40% sont significativement corrélés⁸ avec la première courbe en S, 47% avec la seconde, et 19% avec les deux. Ce dernier score relativement élevé suggère que le changement sémantique est transverse à la diversité des types, ce que laisse déjà entendre la proportion constante des *Aktionsarten*, si bien que des types déjà établis peuvent bénéficier de la nouvelle nuance et voir leur usage augmenter. On peut en tout cas calculer la différence de corrélation entre les deux courbes en S ($rS_2 - rS_1$), et ranger les 203 types suivant ce score pour déterminer leur affinité avec chacune des phases. Nous représentons Figure 5 les trajectoires, en termes de fréquence d'occurrence z-scorées, des 10 types ayant le score le plus élevé et des 10 qui ont le score le plus faible, ce qui permet de visualiser la cohérence de chacun de ces groupes. Les 10 types qui ont le score le plus élevé (donc le plus en faveur de la seconde phase) sont *fuir*, *écouter*, *enfiler*, *téléphoner*, *naître*, *gravir*, *sucer*, *trahir*, *bouffer*, *dormir*, les 10 qui ont le score le plus faible sont *coucher*, *refaire*, *visiter*, *ruiner*, *dresser*, *repasser*, *causer*, *exécuter*, *soigner*, *examiner*. Là encore, le déclin lexical de ces types (en particulier *causer*, majoritairement remplacé par *parler* dans la plupart de ses emplois) peut expliquer ces résultats⁹.

8. Pour déterminer la significativité de la corrélation, nous comparons la valeur p correspondante au seuil traditionnel de 0.05, divisé par le nombre de types (203) pour tenir compte du grand nombre de comparaisons effectuées. Ce critère est donc particulièrement conservatif.

9. Une approche collostructionnelle, fondée sur des scores de spécificité, permettrait d'éviter ces biais, mais ces approches présentent des difficultés méthodologiques, surtout lorsque le nombre d'occurrences est faible : le rang obtenu est alors très largement le résultat de fluctuations statistiques (Gries 2023).

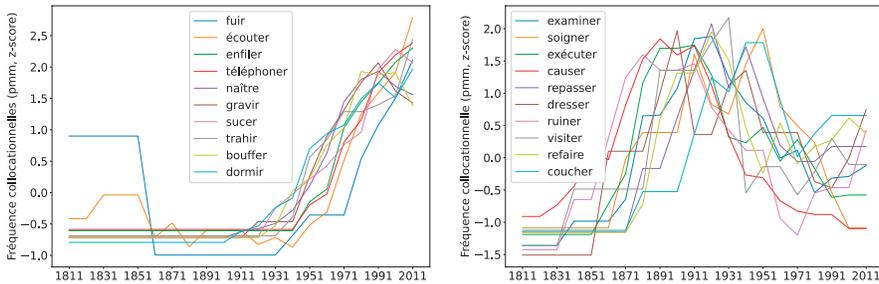


Figure 5. Fréquence d'occurrence z-scorée des 10 verbes présentant le score de différence de corrélation le plus élevé (à gauche) et le plus faible (à droite).

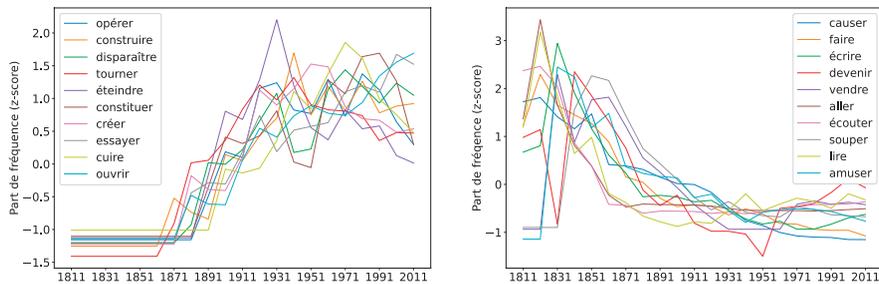


Figure 6. Part de fréquence z-scorée des 10 verbes présentant le score de différence de part de fréquence le plus élevé (à gauche) et le plus faible (à droite).

Nous proposons donc une troisième approche. L'idée principale est que, si un nouveau sens se développe, alors certains types, pré-existants ou non, dont le sémantisme propre s'aligne avec cette nouvelle nuance, verront leur part de fréquence à l'intérieur de la construction augmenter (c'est-à-dire qu'il y aura plus d'occurrences de la construction qui seront des occurrences de ce type ; on a par exemple évoqué que la part de fréquence de *en train de faire* passe de 13% à 9% d'une période à l'autre). Cette approche permet de se placer dans le cadre de référence de la construction elle-même. Pour construire un score mesurant l'association avec les deux phases, on calcule la moyenne de la part de fréquence sur chacune des deux périodes, et on rapporte leur différence à l'écart-type associé à cette part de fréquence sur l'ensemble des deux périodes. Les parts de fréquence z-scorées des 10 types ayant le score le plus élevé (*opérer, construire, disparaître, tourner, éteindre, constituer, créer, essayer, cuire, ouvrir*) et des 10 types ayant le score le plus faible (*causer, faire, écrire, devenir, vendre, aller, écouter, souper, lire, amuser*) sont affichées Figure 6.

D'une manière intéressante, les 10 types voyant leur part de fréquence prendre le plus en importance au sein de la construction suivent un mouvement d'émergence qui s'amorce dès la seconde moitié du XIX^e siècle, ce qui suggère qu'il ne s'agit pas du processus associé à la seconde courbe en S, plus tardive. Ainsi, c'est la seconde méthode, en dépit de ses limitations, qui nous paraît la plus à même d'identifier les types pertinents pour préciser les contours de la nouvelle valeur sémantique dont nous avons fait l'hypothèse, c'est-à-dire ceux dont la combinaison avec la construction est la plus révélatrice de la nouvelle fonction aspectuelle associée.

3.3. Autres associations syntagmatiques

Pour comprendre la seconde courbe, il faut donc se tourner vers d'autres indices. Certaines occurrences montrent que la valeur progressive de *en train de* n'est plus systématique : on trouve par exemple à partir de 1950 des occurrences de *progressivement* (exemple 60), qui indiquent que *en train de* ne suffit plus nécessairement à communiquer la valeur progressive, ou encore *brusquement* (exemple 61), qui paraît incompatible avec le sens progressif :

- (60) Garinati regarde, en-dessous du tablier, l'encombrement des poutrelles métalliques et des câbles, en train de disparaître progressivement à la vue. (A. Robbe-Grillet, *Les Gattes*, 1952)
- (61) Jeanne est auguste. Pas seulement du fessier. De partout. C'est rare à notre époque. Peu de gens aujourd'hui sont augustes. [...] Bien sûr, je ne suis pas en train de découvrir brusquement que Jeanne est auguste : elle l'a toujours été et au fil des ans, ça m'a toujours ou agacée, ou amusée, selon les moments. La nouveauté est qu'à la minute présente, ça m'attendrit. (F. Dorin, *Les Vendanges tardives*, 1997)

En outre, le recours à *toujours*, *encore*, et *déjà* devant *en train de*, dont Do-Hurinville (2007) a déjà souligné l'emploi, émerge essentiellement à partir de la seconde phase, en suivant une courbe en S très proche de celle qui caractérise alors la construction dans son ensemble, ainsi qu'on l'observe dans la Figure 7 (le z-score, qui remplace la valeur de chaque décennie par son écart à la moyenne rapporté à l'écart-type, permet de ramener toutes les courbes à la même échelle, sans altérer leurs variations intrinsèques). De même, on trouve lors de cette seconde phase *en train de* en compagnie d'adverbes modalisateurs, *sûrement*, *sans doute*, et *peut-être*, dont le sens rentre en conflit avec l'idée d'un procès dans lequel l'action s'inscrit. Même si leurs occurrences restent peu nombreuses, leur émergence s'inscrit clairement de manière concomitante au second mouvement d'augmentation de la fréquence (Figure 7).

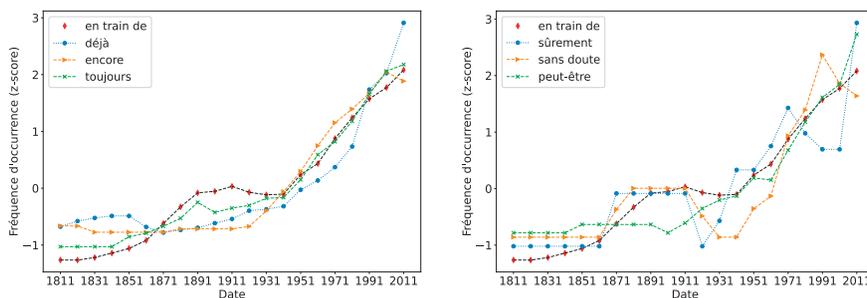


Figure 7. Fréquence d'occurrence z-scoreée de déjà/encore/toujours en train de (à gauche) et de sûrement/sans doute/peut-être en train de (à droite). La fréquence d'occurrence z-scoreée de la construction est donnée en guise de référence.

Enfin, Momma (2023 : 160) signale au vingtième siècle une nette diversification du lexème sur lequel se fixe *en train de* : en plus de *être en train de*, on rencontre désormais (*se*) *voir, trouver, regarder, montrer, croire, rencontrer, décrire, surprendre, découvrir, observer, apercevoir, imaginer, ... en train de*. Là encore, l'émergence de ces emplois se fait de manière concomitante à la seconde courbe en S, comme on peut le voir Figure 8. Par contraste, les emplois où *en train de* s'applique directement à un nom (cf. exemple 2) suivent exactement la même trajectoire que l'ensemble des occurrences (ils n'en représentent pourtant qu'à peine 10%), signe qu'ils ne constituent pas des emplois distincts de l'usage général de cette construction.

Nous sommes donc en présence d'un éventail d'emplois qui se développent de manière concomitante au *xx^e* siècle, et qui ne s'accordent pas nécessairement avec le sens progressif. Nous chercherons dans la section suivante à éclaircir ces emplois.

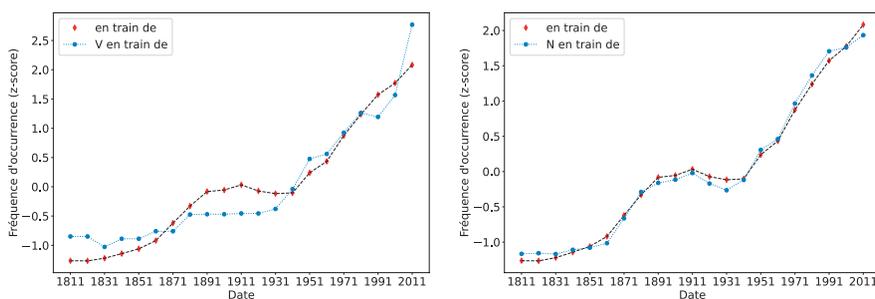


Figure 8. Fréquence d'occurrence z-scoreée de V (excepté être) + en train de (à gauche) et de N + en train de (à droite). La fréquence d'occurrence z-scoreée de la construction donnée en guise de référence.

3.4. Interprétation des résultats

3.4.1. Une désémantisation caractéristique d'un processus de grammaticalisation

Notre hypothèse est la suivante : lorsque le sens progressif de *en train de* se développe au cours de la première phase, la construction garde encore les traces de son sens modal, et se combine essentiellement avec des verbes présupposant un engagement subjectif de la part de l'agent – Lière (2011 : 294) voit ainsi dans l'agentivité l'un des marqueurs de la construction dans ses emplois aspectuels progressifs¹⁰. Les verbes qui, d'après la troisième méthode, sont associés à *en train de* dans la première phase, comme *faire*, *aller*, *lire* et *(s')amuser*, sont ainsi caractéristiques de cet engagement.

Or ce trait sémantique connaît un relâchement, et l'on trouve ainsi, parmi les termes associés à la seconde phase, divers termes de transformation d'état (*cuire*, *bouillir*, *pourrir*), des actions subies par l'agent ou qui présupposent sa passivité (*naître*, *mourir*, *disparaître*, *oublier*, *attendre*, *dormir*, *s'éloigner*), des verbes qui supposent une distanciation évaluative par rapport à l'action (*commettre*, *accomplir*, *essayer*), et des verbes de « corvée » (*laver*, *surveiller*, *balayer*). Cette relaxation du trait d'engagement subjectif peut se décrire comme une désémantisation ou un « blanchiment » sémantique, mais il correspond bien à la prise d'un nouveau domaine sémantique plus vaste qui assure à la construction de nouveaux usages – on rejoint ici la perspective de Himmelmann (2004) décrivant la grammaticalisation comme une extension à de nouveaux contextes. On a par ailleurs déjà mentionné l'affaiblissement de la préférence de *en train de* pour la première personne (Momma 2023 : 97-98), de même que l'émergence de sujets non-animés, dont la part reste à peu près constante au cours des XIX^e, XX^e, et XXI^e siècles (Momma 2023 : 108).

Cependant, cette dé-subjectification ne correspond pas l'émergence d'une nouvelle nuance sémantique : elle est l'aboutissement du processus de grammaticalisation qui a cours au XIX^e siècle. C'est ce que nous avons souligné déjà en remarquant que les types dont la part de fréquence est en moyenne plus élevée lors de la seconde phase que lors de la première (méthode 3) émergeaient dès la fin du XIX^e siècle ; et certains de ces types soulignent bien la perte de l'agentivité (p.ex. *cuire*, *tourner*, *disparaître*). En outre, si l'on se restreint aux seuls essais (Figure 9) où l'engagement subjectif joue un rôle plus faible, on trouve la même courbe en S pour la première phase, seulement décalée plus tardivement dans le temps, et avec un empan similaire (gain de fréquence de 36 occurrences par million de mots contre 41 occurrences par million de mots sur l'ensemble du corpus). Cela signifie

10. De manière curieuse, l'un des exemples cités pour illustrer cette agentivité, tiré de *La Route* de Cormac McCarthy, est : « Les arbres. Ils sont en train de tomber. » Or, les arbres, décrits comme morts, n'ont pas d'agentivité ici ; et d'après nos analyses, *tomber* est caractéristique de la seconde phase, où ce trait s'est déjà relâché.

que la relaxation de la contrainte sémantique d'agentivité/subjectivité liée au sens modal source s'opère dans cette première phase : la première courbe en S est donc bien la signature d'une grammaticalisation à part entière et pleinement aboutie. Des occurrences comme « en train de pleuvoir » sont ainsi attestés dans Google Books dès la toute fin du XIX^e siècle, dans des manuels de linguistique, pour traduire des formes perfectives (depuis l'anglais, le peul, le wallon, le kij). Si *en train de* au XX^e siècle ne traduit plus l'engagement subjectif, il ne s'agit pas d'une nouvelle évolution, mais de l'attestation désormais commune de l'usage progressif désémantisé qui s'est établi à l'issue du processus de grammaticalisation au XIX^e siècle.

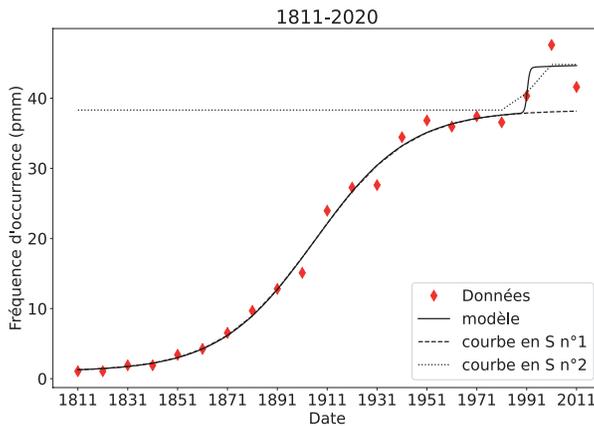


Figure 9. Modélisation par une double courbe en S de la fréquence d'occurrence sur la période 1811-2020 pour le sous-corpus des essais et des écrits scientifiques académiques. Quatre textes ont été exclus (R096, R097, R604 et S855), ceux-ci présentant une forte composante narrative et de ce fait un nombre anormalement élevé d'occurrences relativement au reste du sous-corpus.

3.3.2. L'émergence d'une nouvelle valeur sémantique

Il reste à éclaircir la seconde courbe en S – dont on peut remarquer au passage qu'elle est absente du corpus essais. Pour saisir de quel développement fonctionnel elle pourrait être la signature, nous passerons en revue une série d'occurrences choisies et correspondant chacune aux types dont l'usage avec *en train de* est spécifique de la seconde phase d'après la méthode 2 (cf. Figure 5).

- (62) Je connus donc le bonheur de trotter et même de galoper sur le plateau. Il est si vaste qu'on peut facilement se croire en train de fuir, et même de s'échapper. (J. Giono, *Le Hussard sur le toit*, 1951)

- (63) Des rêves de passion encombraient l'esprit et le cœur de Nadia. Elle se voyait en train de fuir la maison maternelle, elle se voyait emportée en train, en bateau, en traîneau, elle se voyait au fond d'un vallon de verdure entouré de montagnes, dans les bras de Nicolas. (J. d'Ormesson, *Le Vent du soir*, 1985)
- (64) Mais justement on imagine assez bien Pierre Corneille en train d'écouter, curieux, amical, attentif, cette œuvre qui devait lui paraître un peu étrange, un peu attentatoire à sa dignité d'inventeur, cette œuvre qui, somme toute, beaucoup plus qu'un hommage à lui-même, devait lui paraître un hommage à son rival [...]. (R. Brasillach, *Pierre Corneille*, 1938)
- (65) Mais c'est vous qui, il ne l'écoute plus, à l'heure qu'il est, si tout s'était passé normalement, l'hôtesse remettrait hâtivement de l'ordre dans sa chevelure, dans les livres en tout cas c'est ce qu'elle ferait à l'heure qu'il est, et le type serait en train d'enfiler sa chemise dans son pantalon, et ils auraient l'impression très nette que l'avion perd de l'altitude, mal, tellement mal aux oreilles, et le type dirait, dans les livres, en japonais, Comment t'appelles-tu ?, eh bien sûr elle ne comprendrait pas [...]. (J.-L. Benoziolo, *La boîte noire*, 1974)
- (66) Locomotive s'interroge. L'envie est bien là, exorbitante au creux du ventre, mais mettre un pied dans l'appartement, c'est peut-être s'exposer à de graves ennuis... Et si la rousse était en train de téléphoner aux flics, hein ? (J. Vautrin, *Bloody Mary*, 1979)
- (67) Je me disais, en suivant le balancement des branches :les mouvements n'existent jamais tout à fait, ce sont des passages, des intermédiaires entre deux existences, des temps faibles. Je m'apprêtais à les voir sortir du néant, mûrir progressivement, s'épanouir : j'allais enfin surprendre des existences en train de naître. (J.-P. Sartre, *La Nausée*, 1938)
- (68) « Vous parlez de l'homme qui peint, mais je ne vois personne. Pas de silhouette, pas de personnage. Vous arrive-t-il de faire des portraits ? – Pas d'hommes ? Jeune homme, vous vous méprenez, et vous me surprenez. Tout, ici, est l'homme. – Tout ? Je n'en vois qu'un. » Salagnon désigna une petite figure enveloppée d'une robe à plis, difficile à distinguer, en train de graver le premier tiers du sentier, une figure grande comme l'ongle du petit doigt, prête à disparaître derrière une colline. (A. Jenni, *L'Art français de la guerre*, 2011)
- (69) Je risquais pas de me taper Otto. Je sais supporter les privations. Les anciens pouvaient me rabâcher que j'y arriverais, moi aussi, à l'ersatz sexy comme les copains au bout de quelques années, je ne me voyais pas, même dans l'obscurité, en train de lui sucer la pomme à cet artiste déchu. (A. Boudard, *La Cerise*, 1963)
- (70) Qu'est-ce que la vérité dans l'écriture ? Ne suis-je pas en train de trahir l'autobiographie, avec mes petits arrangements nécessaires pour la mise en scène ? (D. Perrut, *Patria o muerte*, 2009)
- (71) – Vous êtes sûre que vous n'avez pas une verrue quelque part ? – Nnnnon – Un grain de beauté alors ? – Si, dit la blonde, confuse et baissant la voix, mais là, et elle touche d'un ongle orange ce qui pointe au nord de son déshabillé. – Épatant, dit Solange. Un gros grain de beauté sur le sein, ça fera une photo du tonnerre, comme un crabe en train de bouffer une méduse comme une mouche sur un gâteau (C. de Rivoire, *Les Sultans*, 1964)
- (72) Maintenant, Hawa est partout, sur les pages des magazines, sur les planches de contact, sur les murs de l'appartement. [...] Hawa debout au-dessus de la mer Méditerranée,

Hawa au milieu de la foule du cours Belsunce, ou bien sur les marches de l'escalier de la gare [...] ; Hawa, en train de marcher, en train de danser, Hawa en train de dormir, Hawa au beau visage couleur de cuivre, au corps long et lisse, qui brille dans la lumière, Hawa au regard d'aigle, aux lourds cheveux noirs qui cascadenent sur ses épaules, ou bien lissés par l'eau de mer comme un casque de galalithe. (J.-M.G Le Clézio, *Désert*, 1980)

Ces occurrences présentent différentes déviations par rapport au sens aspectuel, progressif. Tout d'abord, toutes font référence à une scène qui n'a pas lieu dans la situation de l'énonciation, mais qui est soit imaginée (62, 63, 64, 65, 67, 69) soit dépeinte (68, 72), soit envisagée (66, 70), soit évoquée comme métaphore, comme une image frappante (71). L'occurrence (68) est tout spécialement intéressante, en ce qu'elle se réfère à un tableau statique ; sur ce tableau est représenté un homme « en train de gravir le sentier ». Il s'agit certes d'un procès mais d'un procès fictif, que l'on capture avec une vue instantané qui le résume et l'exprime. L'occurrence (72) décline également toute une série de portraits d'une femme (il s'agit dans le contexte de la narration de photographies de cette femme) : la photographie de Hawa « en train de danser » est donc un instantané, mais qui capture cette dynamique.

On comprend dès lors mieux pourquoi cette seconde phase se caractérise par des verbes de perception visuelle – on vient saisir une scène, prise sur le vif, mais dont l'instantané révèle la dynamique et la logique –, mais également, des adverbes modalisateurs : *en train de* pouvant introduire des scènes imaginées ou supposées, il devient naturel de douter de son actualité. Les résultats de Momma (2023 : 123) soulignent ainsi qu'*en train de* se combine plus fréquemment avec des verbes de modalité dans la période correspondant à notre seconde phase (xx^e et xxi^e siècles), de même qu'avec le conditionnel (Momma 2023 : 121). Si par ailleurs *en train de* ne se réfère plus non à un procès mais à une scène, il est alors logique d'en considérer la récurrence, la répétition, ou le caractère anticipé (*toujours, encore, déjà*). Enfin, il est normal que cette invitation introduite par *en train de* à « faire scène », cette dimension d'hypotopose, soit absente du corpus des essais. Cette dimension scénique explique enfin que des colexèmes à caractère sexuel, comme *sucer* et *baiser*, ou des colexèmes d'action présentant un focus thématique fort, comme *crier* et *pleurer*, soient associés à la seconde phase.

Pour saisir la différence entre le sens progressif et le sens « scénique », on peut comparer ces deux occurrences, tirées du corpus frTenTen20. En (74), il n'y a pas de procès à proprement parler, seulement un instantané (« le voilà ») qui permet de saisir l'essentiel de la scène et d'inviter le locuteur à se la représenter.

(73) Si les responsables nous voient avec une cigarette ou en train de boire un café, on reçoit directement un avertissement. (<https://inegalites.be>)

(74) Deux heures plus tard, le voilà en train de boire un café à une terrasse avant de se rendre au port pour acheter un billet de paquebot pour Marseille. (<https://www.casusno.fr>)

De manière révélatrice, on trouve de nombreuses co-occurrences de *en train de* et de cette idée de *scène* dans Frantext ; nous en avons ici sélectionné trois. L'occurrence (77), en particulier, illustre parfaitement le fonctionnement pragmatique associé à cette nouvelle nuance de sens, lequel consiste, comme nous l'avons évoqué, à inviter le locuteur à élaborer une scène qu'*en train de* vient résumer et saisir.

- (75) Je me suis retrouvé à la maison aussitôt, j'ai vécu cette scène : mon père tous les jours en train de lire, à l'heure des repas, trois, quatre pages de mon livre, le voyage en virginie, de longues, interminables phrases avec les virgules des crochets où étrangler le temps. (Y. Berger, *Le Sud*, 1962)
- (76) La démonstration apparaît encore plus forte lorsqu'elle est volontairement artificielle : par la porte ouverte d'une boutique d'armes, on aperçoit dans la rue deux hommes en train de ferrailer : la scène est peu probable, logique cependant si l'on veut montrer le terme ultime de l'opération (sujet de la planche), qui est le fourbissage [...]. (R. Barthes, *Nouveaux essais critiques : Les planches de l'«Encyclopédie»*, 1964)
- (77) – Pas réglo.
 – Pas réglo du tout. Dinh, imagine qu'un salopard m'ait lancé tout habillé dans un lac de merde.
 – Cela arrive.
 – Et que je sois en train de m'enfoncer dedans. Tu visualises la scène ?
 (F. Vargas, *Un lieu incertain*, 2008)

Remarquons enfin que la valeur aspectuelle et la dimension scénique ne s'opposent pas nécessairement, et peuvent fonctionner de manière simultanée ; c'est en particulier le cas lorsque *en train de* sert de cadratif à une action plus significative, comme dans l'occurrence suivante :

- (78) J'étais en train de voler des fruits dans un jardin, des pommes excellentes et nacrées, quand le voile se déchira et qu'à ma grande terreur, une voix que j'avais complètement oubliée résonna derrière moi. Je me suis retourné lentement, la bouche pleine, fâché de ma saleté qui était grande (je venais de traverser des ronciers et je m'étais battu avec un chien-loup). Il fallait croire que c'était elle. (R. Nimier, *Le Hussard bleu*, 1950)

D'une part, l'enchaînement action en cours/interruption (*être en train de X quand Y*) renforce la valeur progressive, l'idée d'un procès en cours ; d'autre part, l'action de voler des fruits dans un jardin se distribue en une multitude de petits actes et *en train de* ne se réfère à aucun spécifiquement, ce qui ne permet pas de situer à quel moment de la séquence a lieu l'interruption (on en déduit plus tard que c'est au moment où il déguste une des pommes). En outre, la spécification de « pommes excellentes et nacrées », dont il n'est nul besoin pour préciser l'action que vient interrompre l'arrivée de la jeune femme, se fait au service de la scène, et vise à en enrichir la représentation par le lecteur. On peut donc considérer qu'il existe un continuum d'emplois entre le progressif et le scénique, qui peuvent dès lors être activés simultanément ; mais en

même temps les deux sens sont suffisamment différents pour pouvoir être employés indépendamment, l'un à l'exclusion de l'autre. C'est en particulier le cas lorsque *en train de* se réfère à quelque chose qui n'a pas encore lieu, mais qui est imminent, et dont on suggère donc l'éventualité en appelant à la visualiser, à l'envisager – on peut ainsi rattacher les emplois d'inadvertance et mise en garde à cette nuance sémantique. L'occurrence (79) illustre ainsi le caractère virtuel mais imminent de l'action introduite par *en train de*, en particulier à travers la forme *c'est en train de*, spécifique de cette nuance et exclusive à la seconde phase :

- (79) à propos, lui dit Zazie, je crois que c'est en train de me revenir la question que je voulais te poser. (R. Queneau, *Zazie dans le métro*, 1959)

Terminons par l'exemple cité en ouverture d'article : dans cette phrase, l'auteur effectue une sorte de prosopopée, élabore une sorte de mise en scène concentrée d'un dépassement des savoirs, d'un renouvellement général des connaissances – il s'agit bien, et pleinement, de l'aspect scénique ici dégagé.

Conclusion

Dans cet article, nous avons cherché à éclairer la structure sémantique complexe de la préposition complexe *en train de*, tant sur le plan diachronique que sur le plan synchronique. Nos analyses s'enracinent dans l'idée, héritée des formalismes de la sémantique vectorielle, que le sens forme un domaine structuré que viennent « peupler » les occurrences. Si chaque occurrence constitue, dans cette perspective, sa propre nuance de sens, elles s'organisent collectivement en clusters qui permettent de faire émerger des régions distinctes les unes des autres, correspondant à un paysage sémantique multipolaire, qui permet d'asseoir empiriquement l'idée de polysémie. Cependant, les techniques actuelles permettent généralement de décrire la structure sémantique à différents niveaux de granularité dans la représentation, mais ne définissent pas de niveau privilégié qui permettrait de répondre à la question « Quand y a-t-il plusieurs sens ? » (cf. Kleiber 2005).

Pour y répondre, nous avons adopté une perspective diachronique : sous l'hypothèse que l'extension d'une forme à un domaine sémantique induise une augmentation de sa fréquence d'occurrence, et que cette augmentation obéisse au motif bien établi de la courbe en S, alors repérer différentes courbes en S s'additionnant les unes aux autres dans le développement diachronique d'une forme permettrait de repérer les différentes nuances de sens qui se trouvent associées à cette forme au cours de ses vicissitudes historiques.

Pour pouvoir l'appliquer, il était néanmoins nécessaire de bien repérer le moment où *en train de* commence à être utilisé dans son sens progressif – or cette

question fait débat, certains le considérant comme datant du xvi^e siècle au moins, lorsque d'autres y voient le résultat d'une grammaticalisation à partir d'un sens modal. Sur la base de l'examen qualitatif d'un vaste éventail d'occurrences, nous avons avancé l'idée que, si *en train de* (ou plutôt la séquence *en + train*) a toujours été associé à des fonctions aspectuelles (notamment inchoatives et imminentielles), pas nécessairement bien arrêtées et émergeant souvent de critères contextuels, la construction en tant que telle émerge bien au xvii^e siècle avec un sens modal, le sens aspectuel progressif ne se développant réellement qu'au xix^e siècle, ainsi que le considérait déjà Gougenheim (1929). Ce processus obéit par ailleurs au parcours habituel d'un processus de grammaticalisation : basculement sémantique à partir d'un sens source (ici un sens modal de disposition), forte augmentation de fréquence suivant une courbe en S, désémantisation/extension des contextes d'emploi à mesure que la valeur d'implication subjective perd son importance. Ce sens modal s'affaiblit en outre dès la fin du xviii^e siècle et disparaît à l'issue du xix^e siècle.

Nous avons ensuite appliqué le motif de la courbe en S au développement de la construction à partir de l'apparition du sens progressif. Il apparaît alors qu'à partir des premières décennies du xx^e siècle, une seconde phase de l'évolution a lieu, qui conduit à doubler la fréquence d'emploi de la préposition, là encore suivant une courbe en S. Nous voyons dans ce phénomène la traduction d'une extension sémantique ultérieure, à laquelle seraient rattachés les emplois de « discordance » (Leeman 2012) ou « antitéléonomiques » (Franckel 1989). Cette extension sémantique ne peut pas être facilement établie à partir des relations de colexification entre la préposition et le verbe qu'elle introduit, car les valeurs aspectuelles mises en jeu dans l'une et l'autre nuance peuvent s'appliquer aux mêmes prédicats. Sur la base de différents indices quantitatifs (développement d'autres verbes que *être* en support de la construction, corrélation de la trajectoire de certains lexèmes avec la seconde courbe en S, combinaison avec les adverbes temporels *toujours*, *encore*, *déjà*), nous avons indiqué que la seconde phase se démarque effectivement de la première en termes d'emplois inédits, qui restent à caractériser pleinement.

Dans cette veine, nous avançons l'hypothèse que cette nuance sémantique additionnelle permet de souligner l'aspect « scénique » de l'action, posée en tant que focus thématique de la représentation ainsi convoquée, et qu'elle se double en outre d'une fonction pragmatique, invitant le locuteur ou le lecteur à visualiser la scène à partir d'un trait dynamique essentiel que vient saisir *en train de*. On notera que tous les emplois de *en train de* rencontrés mettent l'accent sur cette dimension dynamique, que Do-Hurinville (2007) regardait comme l'invariant sémantique de cette construction.

La diversité des emplois de *en train de*, ce développement diachronique en deux phases qui viennent se superposer l'une à l'autre, suggèrent donc que la préposition est fondamentalement polysémique, et présente au moins deux « domaines » de sens, qui coexistent en synchronie et demeurent l'un

et l'autre extrêmement productifs, assurant la fortune de cette préposition complexe et atypique¹¹.

Références

- ANSCOMBRE J.-C. (2007). Les indicateurs aspectuels de déroulement processif : 'en cours de', 'en passe de', 'en train de', 'en voie de'. *Cahiers de lexicologie* 90, 41-74.
- ATILF. (1998-2023). *Base textuelle Frantext (En ligne)*. ATILF-CNRS & Université de Lorraine. <https://www.frantext.fr>
- BERTINETTO P.M. (2000). The progressive in Romance, as compared with English. *Empirical approaches to language typology* 6, 559-604.
- BESSE, H. (1970). Problèmes de sens dans l'enseignement d'une langue étrangère. *Langue française* 8, 62-77.
- BLYTHE R.A. & CROFT W. (2012). S-curves and the mechanisms of propagation in language change. *Language* 88 (2), 269-304.
- BOTTINEAU D. (2014). La spécificité du « progressif » catalan : une approche contrastive (anglais, français, gallo, italien, castillan). In : M. Pujol Berché (éd.), *Recherches sur la langue catalane: actes du Colloque international, Nanterre, 4 et 5 novembre 2010*. Limoges : Lambert-Lucas, Limoges, 151-174.
- BRUGMAN C. & LAKOFF G. (1988). Cognitive topology and lexical networks. In : S. L. Small, G. W. Cottrell, M. K. Tannenhaus (eds), *Lexical ambiguity resolution*. San Mateo : Morgan Kauffmann, 477-508.
- DE MULDER W. (2019). *En voie de* : du trajet spatial à l'aspect. *Revue Romane. Langue et littérature. International Journal of Romance Languages and Literatures* 54 (1), 39-61.
- DE WIT A. & PATARD A. (2013). Modality, aspect and the progressive: The semantics of the present progressive in French in comparison with English. *Languages in contrast* 13(1), 113-132.
- DESAGULIER G. (2022). Changes in the midst of a construction network: a diachronic construction grammar approach to complex prepositions denoting internal location. *Cognitive Linguistics* 33(2), 339-386.
- DEVLIN J., CHANG M.W., LEE K. & TOUTANOVA K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In : *Proceedings of NAACL-HLT 2019*, Minneapolis : Association for Computational Linguistics, 4171-4186.

11. L'auteur remercie Lydia Momma pour lui avoir communiqué sa thèse de doctorat, ainsi que les deux relecteurs anonymes, dont les remarques nombreuses et fécondes ont permis de donner à cet article un caractère nettement plus abouti. L'auteur bénéficie en outre d'un financement de l'université de Gand, BOF.PDO.2022.0001.01.

- DO-HURINVILLE D.T. (2007). Etude sémantique et syntaxique de *être en train de*. *L'Information grammaticale* 113, 32-39.
- FELTGEN Q. (2022). Ce que les variations de fréquence nous apprennent des changements linguistiques : le cas de la construction *en plein* N. *Langue française* 215(3), 61-80.
- FELTGEN Q. (2024). Is language change chiefly a social diffusion affair? The role of entrenchment in frequency increase and in the emergence of complex structural patterns. *Frontiers in Complex Systems*, 2, 1327425.
- FELTGEN Q., FAGARD B. & NADAL J.-P. (2017). Frequency patterns of semantic change: corpus-based evidence of a near-critical dynamics in language change. *Royal Society open science* 4(11), 170830.
- FONTEYN L. (2021). Varying Abstractions: a conceptual vs. distributional view on prepositional polysemy. *Glossa: a journal of general linguistics* 6(1): 90.
- FRANCKEL J.-J. (1989). *Étude de quelques marqueurs aspectuels du français*. Paris : Droz.
- GOSELIN L. (2011). L'aspect de phase en français: le rôle des périphrases verbales. *Journal of French Language Studies* 21(2), 149-171.
- GOUGENHEIM G. (1929). *Étude sur les périphrases verbales de la langue française*. Paris : Les Belles Lettres.
- GRIES S.T. (2023). Overhauling Collostructional Analysis: Towards More Descriptive Simplicity and More Explanatory Adequacy. *Cognitive Semantics* 9(3), 351-386.
- HAMMA B. (2005). La préposition *par*, génératrice de polylexicalité ? *Linx. Revue des linguistes de l'université Paris X Nanterre* 53, 87-102.
- HEINE B. (2002). On the role of context in grammaticalization. In: I. Wischer, G. Diewald (eds). *New Reflections on Grammaticalization*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 83-102.
- HEYLEN K., WIELFAERT T., SPEELMAN D. & GEERAERTS D. (2015). Monitoring polysemy: Word space models as a tool for large-scale lexical semantic analysis. *Lingua* 157, 153-172.
- HILPERT M. (2016). Change in modal meanings: Another look at the shifting collocates of *may*. *Constructions and Frames* 8(1), 66-85.
- HIMMELMANN N.P. (2004). Lexicalization and grammaticization: Opposite or orthogonal?. In : W. Bisang, N.P. Himmelmann, B. Wiemer (eds), *What makes Grammaticalization? A Look from its Fringes and its Components*. Berlin/New York: Mouton de Gruyter, 21-42.
- JI H. & PLOUX S. (2003). Automatic Contexonym Organizing Model (ACOM). *Proceedings of the Annual Meeting of the Cognitive Science Society* 25, 622-627.
- De JOINVILLE J. (1998). *Vie de Saint Louis*. Paris : Classique Garnier.
- KLEIBER G. (2005). Quand y a-t-il plusieurs sens ? Le critère référentiel en question. In : O. Soutet (éd.), *La polysémie*. Paris : Presses de l'Université Paris-Sorbonne, 51-73.

- KLEIBER G. (2008). Petit essai pour montrer que la polysémie n'est pas un sens interdit. In : J. Durand, B. Habert & B. Laks (éds), *Congrès Mondial de Linguistique Française – CMLF'08*, Paris : Institut de Linguistique Française, 87-101.
- KROCH A. (1989). Reflexes of grammar in patterns of language change. *Language variation and change* 1(3), 199-244.
- LACHAUX F. (2005). La périphrase *être en train de*, perspective interlinguale (anglais-français): une modalisation de l'aspect. In : N. Le Querler, H. Bat-Zeev Shyldkrot (eds), *Les périphrases verbales*. Amsterdam : John Benjamins Publishing Company, 119-142.
- LANGACKER R.W. (1987). Mouvement abstrait. *Langue française* 76, 59-76.
- LAROUSSE P. (1876). *Grand dictionnaire universel du XIXe siècle*, T. 15. Paris : Administration du grand dictionnaire universel.
- LAUWERS P. & TOBBACK E. (2010). Les verbes attributifs : inventaire(s) et statut(s). *Langages* 179-180, 79-113.
- LEEMAN D. (2012). Contribution à la définition de l'identité de la périphrase *être en train de*. *La Tribune internationale des langues vivantes*, N° spécial, juillet 2012, 133-138.
- LEEMAN D. & VAGUER C. (2014). La préposition peut-elle être prédicative ? Le cas de la préposition "en". *Verbum* 36, 397-419.
- LIÈRE A. (2011). *Entre lexique et grammaire : Les périphrases verbales du français*. Thèse de doctorat, Université du Littoral – Côte d'Opale, Dunkerque.
- MERLE J.-M. (2008). Prépositions et aspect. *L'Information grammaticale* 117, 52-56.
- MIKOLOV T., SUTSKEVER I., CHEN K., CORRADO G.S. & DEAN J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems* 26, 3136-3144.
- MOMMA L. (2023). *Progressive Verbalperiphrasen: Eine korpusbasierte diachrone Untersuchung zur Grammatikalisierung von frz. "être en train de faire"*. Thèse de doctorat, Eberhard Karls Universität, Tübingen.
- MONTES M. & GEERAERTS D. (2022). How vector space models disambiguate adjectives: A perilous but valid enterprise. *Yearbook of the German Cognitive Linguistics Association*, 10(1), 7-32.
- MORTIER L. (2005). Les périphrases aspectuelles « progressives » en français et en néerlandais : présentation et voies de grammaticalisation. In : N. Le Querler, H. Bat-Zeev Shyldkrot (eds), *Les périphrases verbales*. Amsterdam : John Benjamins Publishing Company, 83-102.
- MORTIER L. (2008). An analysis of progressive aspect in French and Dutch in terms of variation and specialization. *Languages in contrast* 8(1), 1-20.
- NORDE M. & BEIJERING K. (2014). Facing interfaces. A clustering approach to grammaticalization and related changes. *Folia Linguistica* 48(2), 385-424.
- PEREK F. (2016). Using distributional semantics to study syntactic productivity in diachrony: A case study. *Linguistics* 54(1), 149-188.

- PEREK F. (2018). Recent change in the productivity and schematicity of the way-construction: A distributional semantic analysis. *Corpus Linguistics and Linguistic Theory* 14(1), 65-97.
- PEREK F. & HILPERT M. (2017). A distributional semantic approach to the periodization of change in the productivity of constructions. *International Journal of Corpus Linguistics* 22(4), 490-520.
- PLOUX S. & VICTORRI B. (1998). Construction d'espaces sémantiques à l'aide de dictionnaires de synonymes. *TAL* 39, 161-182.
- PLOUX S., BOUSSIDAN A. & JI H. (2010). The semantic atlas: an interactive model of lexical representation. In: N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odiijk, S. Piperidis, M. Rosner & D. Tapias (eds), *Proceedings of the seventh conference of International Language Resources and Evaluation*. Valetta: European Language Resources Association (ELRA), 1-5.
- PUSCH C.D. (2003). La grammaticalisation de l'aspectualité : Les périphrases à valeur progressive en français. *Verbum* 25(4), 495-508.
- SALOFF-COSTE M. (2019). Les écosystèmes, forces motrices de l'innovation. Expériences internationales. In : B. Laperche, M. Lima, E. Seulliet & B. Trousse (éds), *Les écosystèmes d'innovation*. Paris : L'Harmattan, 47-68.
- VENDLER Z. (1957). Verbs and times. *The philosophical review* 66(2), 143-160.
- VETTERS C. (1989). Grammaticalité au passé récent. *Linguisticae investigationes* 13(2), 369-386.

n° 236

Vient de paraître

Varia

La composition de ce numéro a été confiée à Catherine Schnecker.

L'acquisition de l'aspect accompli par les étudiants jordaniens : principe des sous-ensembles, transfert ou facteurs multiples ? – Authoul Abdul Hay, Alaualdeen Algharibeh

De l'hésitation à la correction dans les aphasies non fluentes : étude exploratoire du complexe « auto-rectifications-gestes-pauses – Corrado Bellifemine

Pronoms clitiques : une variation référentielle – Alain Berrendonner

Introducing Regionalisms in General French Language Dictionaries: A Case Study of the *Dictionary of the French Academy* – Anica Bojic

L'IA dispose-t-elle d'une compétence idiomatique ? Étude de cas de la traduction automatique des expressions figées françaises en anglais et en chinois – Xiaoyao Chi

La satisfaction des besoins psychologiques fondamentaux et la motivation d'apprendre le français des adolescents – Madis Olsper, Merylyn Meristo

Temporal Regression and Tense: The case of the French Imparfait and Passé simple – Adeline Patard

PARUTION : décembre 2024

172 pages • 18 €

EAN : 978-2-200-93537-5

