

J. L. LEBRAVE

La fréquence des nominalisations en allemand

Mathématiques et sciences humaines, tome 52 (1975), p. 35-54

http://www.numdam.org/item?id=MSH_1975__52__35_0

© Centre d'analyse et de mathématiques sociales de l'EHESS, 1975, tous droits réservés.

L'accès aux archives de la revue « Mathématiques et sciences humaines » (<http://msh.revues.org/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

LA FREQUENCE DES NOMINALISATIONS EN ALLEMAND

J.L. LEBRAVE*

Les nominalisations sont-elles caractéristiques de certains styles? Telle est la question à laquelle ce travail se proposait initialement de répondre, en appliquant une définition linguistique rigoureuse à l'étude d'un phénomène stylistique qui passe traditionnellement pour bien connu, le style nominal. Nous ne nous étendrons pas ici sur les problèmes spécifiquement linguistiques posés par la nominalisation et son rapport avec le style nominal. Les traités de stylistique allemande définissent en général le style nominal par l'abondance des "substantifs", et l'apport de la linguistique consiste d'abord à quitter le terrain des "classes de mots" (verbe, substantif, etc.) pour celui d'unités linguistiques mieux définies¹. On peut ainsi distinguer à l'intérieur du style nominal le procédé qui consiste à remplacer une expression verbale par une expression nominale (et dans ce cas il s'agit donc de nominalisations), et celui par lequel un auteur évite systématiquement d'employer des verbes (c'est par exemple le cas pour certains expressionnistes allemands). Nous ne nous intéresserons ici qu'au premier de ces deux procédés, et nous admettrons qu'on peut donner des nominalisations une définition opératoire, applicable à l'étude d'un corpus.

* Centre d'Histoire et d'Analyse des Manuscrits Modernes (C.N.R.S.)

1) On voit mal quel est l'intérêt de ranger sous la même rubrique les deux phrases suivantes, comme le font par exemple W.FLEISCHER et G.MICHEL (*Stilistik der deutschen Gegenwartssprache*, Leipzig 1975, pp.230-233): "Die krankhafte Erhöhung des Blutdrucks beruht zum Teil auf einer ständigen Erhöhung des Gefäßwiderstandes der Masse der kleinen Arterien" et "Szenen, Auftritte, ein Monolog, ein behördlicher Einspruch, die Stiftung eines Verlöbnisses, ein scharfer Dialog, vorher aber: der Auszug eines jungen Menschen, vom Dorf nach der Stadt, eines Witwenkindes, der Abschied am Zaun."

Malheureusement, l'apport de la linguistique cesse dès qu'on aborde l'étude stylistique proprement dite, où l'on ne dispose que d'approches relativement traditionnelles, toutes fondées en dernier ressort sur l'intuition du chercheur et sur sa connaissance des textes étudiés. Lorsqu'on essaie d'appliquer ces intuitions à un corpus précis, on aboutit rapidement à une impasse: selon la conception qu'on se fait de tel auteur, de sa place dans le mouvement des idées, de son importance dans l'évolution de la langue, on pourra trouver des arguments convaincants susceptibles de confirmer les thèses les plus contradictoires, sans que rien permette en fait de choisir entre les deux. Par exemple, l'étude linguistique montre que les nominalisations sont caractéristiques de la langue scientifique et technique¹. Mais qu'en est-il du style de Heine critique littéraire? Selon l'importance qu'on donne aux préoccupations philosophiques et politiques de Heine, on cherchera dans les textes - et on trouvera - les marques d'une subjectivité impénitente inapte à l'objectivation qu'implique l'emploi des nominalisations, ou au contraire on s'efforcera de montrer qu'il introduit dans la critique littéraire une dimension scientifique, et donc que son style se caractérise par un usage marqué des nominalisations. Pour sortir de cette impasse, il est nécessaire de revenir sur les mécanismes intuitifs d'analyse du "style" en cherchant à quelles hypothèses implicites ils renvoient. On constate en particulier que ces mécanismes mettent en jeu des hypothèses probabilistes au nom desquelles on juge de la pertinence stylistique d'un phénomène, qu'il s'agisse de sa "fréquence" ou de sa valeur.

Nous montrerons comment on peut formuler explicitement ces hypothèses, en cherchant à vérifier leur pertinence par le recours à des procédures d'ajustement statistique. On verra dans quelle mesure la distribution des nominalisations dans une oeuvre, littéraire ou non, est réductible à un processus aléatoire (processus de Poisson), et on essaiera brièvement d'envisager les conséquences de ce résultat sur l'analyse stylistique et sur la pertinence stylistique des nominalisations.

° °

°

1) Cf. K. MOESLEIN, Der Nebensatz und sein nominales Aequivalent in der wissenschaftlich-technischen Literatur des XIX. und XX. Jahrhunderts. Habilitationsschrift. Leipzig 1968.

Du point de vue linguistique, on peut définir les nominalisations par leur fonction dans la phrase, par leur forme, ou par leur spécificité sémantique.

Par la fonction: les nominalisations font partie des éléments qui, dans une phrase, peuvent se substituer à un groupe nominal. Une telle définition est de peu d'intérêt dans le cadre du travail entrepris ici, puisqu'elle regroupe sous le terme de nominalisations, outre des expressions nominales, certaines "subordonnées" comme les dass-Sätze, des groupes infinitivaux, etc.¹.

Par la forme: parmi l'ensemble des groupes nominaux, il en existe qui présentent certains traits spécifiques qui les distinguent de tous les autres:

- la base comporte un suffixe caractéristique (das -en, die -ung, d- -Ø, etc.);
- on peut établir une correspondance terme à terme régulière entre ces groupes nominaux et des groupes verbaux comportant les mêmes éléments;
- ces groupes nominaux sont par certains aspects de "pseudo-groupes nominaux": une forte proportion d'entre eux est toujours au "défini-singulier" et ne supporte donc pas le jeu normal des oppositions de catégories, ce qui laisse supposer qu'il ne s'agit pas de groupes nominaux quelconques (la forme est ici l'indice d'un contenu sémantique spécifique et non nominal)².

Par le sens: la correspondance entre nominalisation et groupe verbal n'est pas seulement formelle, mais fondamentalement sémantique: on retrouve dans la nominalisation et dans le groupe verbal correspondant le même "complexe connexionnel" de signifiés³; dans le groupe verbal, ce complexe est mis en relation avec les différentes catégories verbales (temps, mode, modalisation); dans le groupe nominal, il est associé à un élément nominal et actualisé par les catégories nominales comme n'importe quel groupe nominal⁴.

1) C'est en ces termes qu'on peut, en première approximation, résumer la définition de la nominalisation dans le cadre de la grammaire générative: c'est "l'apparition de S sous N", autrement dit la sous-classe des éléments susceptibles de commuter avec des groupes nominaux sans en être eux-mêmes. Cf. par exemple R.B. LEES, The Grammar of English Nominalizations. Bloomington. 1960.

2) Pour une étude linguistique détaillée des nominalisations dans le cadre de cette définition, cf. C. MILNER, Des groupes nominaux en rapport exclusif avec des groupes verbaux. Thèse de doctorat de 3e cycle. Paris 1967 (Dactyl.)

3) Pour une justification de l'analyse des "groupes spécifiques" en complexes connexionnels de signifiés, cf. J. FOURQUET, Prolegomena zu einer deutschen Grammatik, Düsseldorf 1970.

4) Cf. J.L. LEBRAVE, La fréquence des nominalisations. Thèse de doctorat de 3e cycle. Paris 1973 (Dactyl.)

Dans le travail dont on donne ici les résultats, on a commencé par utiliser la définition par la forme, qui semblait convenir particulièrement bien à l'étude d'un corpus, dans la mesure où elle mettait en jeu le seul signifiant. Mais le corpus lui-même en a montré l'insuffisance. D'une part, la présence des trois traits définissants ne permet pas de conclure que le groupe nominal en question est bien une nominalisation: il subsiste un assez grand nombre de groupes nominaux où seul l'examen du sens permet de décider s'il y a ou non correspondance avec un groupe verbal¹. La définition est donc, sur ce point, trop large. D'autre part, l'un de ces traits (la "neutralisation" des oppositions de catégories) n'est pas en fait un trait formel, mais un trait sémantique, voire ontologique: on admet que le référent d'une nominalisation ne peut qu'être unique. Or les textes étudiés contiennent un certain nombre de groupes nominaux qui obligent à repenser le problème du statut des catégories dans la nominalisation, étant donné la nature par ailleurs évidemment "verbale" de ces groupes.

D'où l'élaboration de la troisième définition, qui permet de distinguer trois types d'unités différentes:

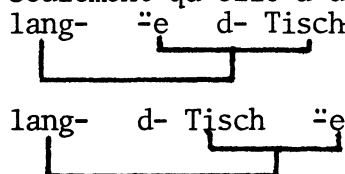
a) les "BN^V": le complexe de signifiés verbal s'associe à un sème nominal pour constituer une base nominale complexe susceptible de recevoir des membres comme n'importe quelle base, et compatible avec toutes les catégories nominales. Le suffixe est par exemple die -ung, das -en, d- -Ø, etc. On aura ainsi:

die Bewältigung der Materie durch den Geist
qui s'analyse de la façon suivante:

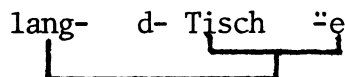
1) Un exemple simple montrera ce qu'on entend ici par examen du sens. Le groupe nominal die Länge des Tisches possède les traits suivants:

- présence d'un suffixe caractéristique (die -e);
- possibilité d'établir une correspondance terme à terme avec le groupe verbal der Tisch ist lang selon le schéma suivant: au groupe nominal au nominatif du groupe verbal correspond le même groupe nominal au génitif dans la nominalisation; à la base du groupe verbal (lang s-) correspond le lexème lang- dans le groupe nominal;
- le groupe est au défini-singulier.

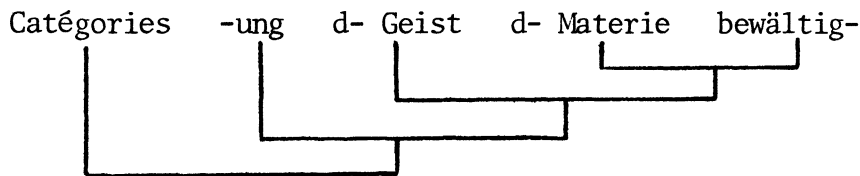
Mais dans son contexte normal d'utilisation (par exemple ich messe die Länge des Tisches), il est évident qu'il ne s'agit pas de dire que la table est longue, mais seulement qu'elle a une certaine longueur. L'analyse du signifié est donc:



et non pas:



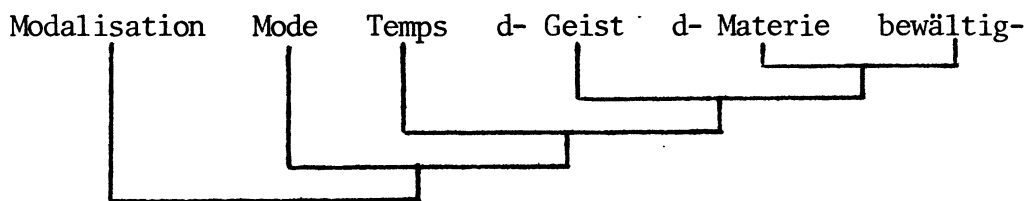
Examiner le sens revient à vérifier si la correspondance avec un groupe verbal est légitime.



On y retrouve le complexe de signifiés commun aux groupes verbaux suivants:

der Geist bewältigt (bewältigte, wird ... bewältigen, etc.) die Materie
 der Geist würde die Materie bewältigen
 Vielleicht bewältigt der Geist die Materie
 etc.

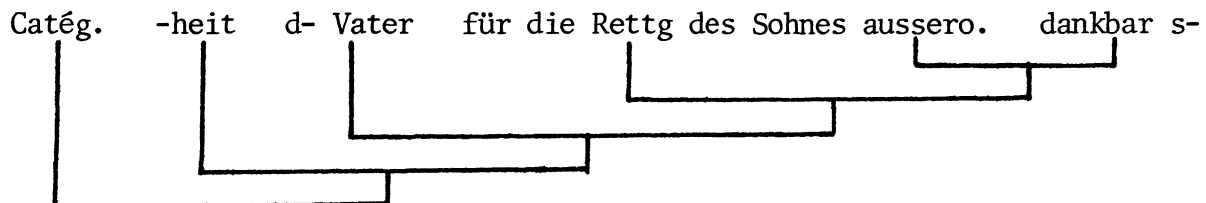
dont l'analyse commune serait:



Ce sont les nominalisations à proprement parler.

b) les "GN^Qs": leur caractère verbal est beaucoup moins net que celui des BN^V, et le complexe de signifiés verbal n'est reconstituable que dans certains cas. Lorsqu'on peut établir la correspondance, elle est entre un complexe verbal dont la base est l'association d'un qualificatif et de sein et le même complexe associé à un suffixe spécifique (die -heit, die -e, etc.). Ainsi:

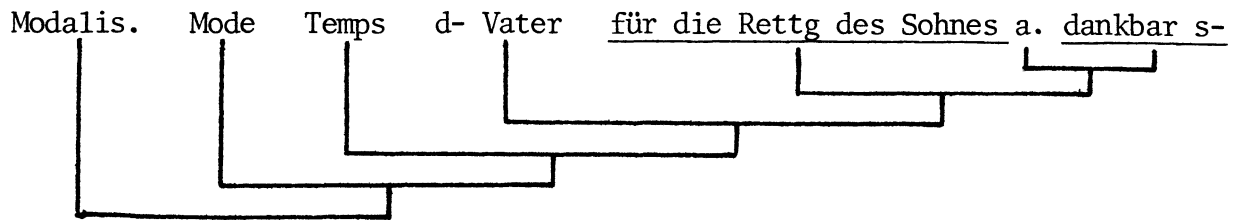
die ausserordentliche Dankbarkeit des Vaters für die Rettung des Sohnes
 peut s'analyser comme:



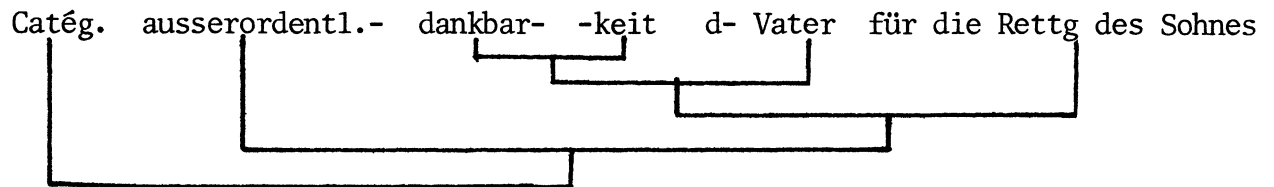
On y retrouve le complexe de signifiés commun aux groupes verbaux suivants:

der Vater ist (war, wird ... sein, etc) für die Rettung des Sohnes
 ausserordenltich dankbar
 der Vater wäre für die Rettung des Kindes ausserordentlich dankbar
 vielleicht ist der Vater für die Rettung des Sohnes ausserordentlich
 dankbar
 etc.

dont l'analyse commune serait:



Mais, même dans cet exemple précis, rien n'empêche de faire une analyse différente, dans laquelle il n'y a pas de correspondance avec un complexe verbal, mais "lexicalisation"¹ du qualificatif, d'où résulte un "abstrait" qualificatif:



c) les "GN^Qdas": le groupe verbal dont on extrait le complexe verbal correspondant à ce type de nominalisation est de la forme:

was + an + groupe qualificatif + ist .

Ainsi, das Schöne daran correspond au groupe verbal:

was daran schön ist .

Toutefois, à la différence des deux cas précédents, le groupe verbal et le groupe nominal sont directement substituables l'un à l'autre. Il n'y a donc pas nominalisation au même titre que, par exemple, dans le cas des BN^V.

On verra que ces différences dans le statut linguistique apparaissent dans le comportement statistique de ces différentes unités.

L'étude linguistique permet de cerner avec plus de rigueur le phénomène étudié, en donnant des nominalisations une définition opératoire applicable à n'importe quel corpus. Mais elle ne donne pas d'indications sur ce qu'on entend par un usage marqué des nominalisations. Si on prend l'exemple de Heine, un lecteur compétent (c'est-à-dire qui connaisse bien l'oeuvre de Heine, la littérature du XIXe siècle, les grands traits de l'évolution de

1) Par lexicalisation, on entend ici la "dégrammaticalisation" des connexions à l'intérieur d'un syntagme, au cours de laquelle ce syntagme cesse d'avoir un sens déductible de celui de ses composants pour acquérir celui d'un lexème simple. C'est ici l'association du qualificatif et du suffixe qui est "lexicalisée". Cf. par exemple Ch. BALLY, Linguistique générale et linguistique française, Berne 1944, pp. 146 et suiv.

l'allemand entre la fin du XVIIIe siècle et nos jours, etc.) peut dire avec quelque apparence de raison que les nominalisations sont caractéristiques de ses textes critiques. Ce jugement peut se justifier de deux façons différentes:

1) Ou bien Heine utilise des expressions nominales là où ses prédécesseurs ou ses contemporains auraient employé des expressions verbales; de ce fait, les textes de Heine sont plus riches en nominalisations que ne le laisseraient prévoir les normes d'utilisation du style nominal dans l'allemand littéraire du XIXe siècle. L'usage marqué des nominalisations se traduit donc par une fréquence anormale des nominalisations. C'est d'ailleurs en ces termes que le style nominal (et son corrélat, le style "verbal") est défini dans les traités de stylistique. Voici par exemple comment W. Fleischer et G. Michel classent les textes selon leur caractère "nominal" ou "verbal": "Derartige Texte sind durch einen über den normalen 'Erwartungswert' hinausgehenden Anteil nominaler (...) oder andererseits verbaler Elemente gekennzeichnet"¹. Malheureusement, les traités de stylistique restent muets sur ce qu'est une "proportion d'éléments supérieure à ce qu'on attendrait normalement"². Or toute tentative d'application de ce type de classification à des textes précis est immédiatement confrontée à la nécessité de définir la "norme d'attente" dont parlent W. Fleischer et G. Michel. Ceci suppose qu'on sache compter combien un texte contient d'éléments nominaux, et surtout qu'on soit en mesure de comparer la fréquence de ces éléments dans différents textes. Pour reprendre l'exemple de Heine, on est amené à vérifier que les textes de Heine contiennent bien un plus grand nombre de nominalisations que ceux des écrivains de la même époque traitant de sujets équivalents. Ceci revient à comparer différents textes quant à la proportion de nominalisations qu'ils contiennent. Or un dénombrement, même rapide, montre immédiatement qu'on ne peut se contenter de ce calcul de proportions. Celui-

1) W. FLEISCHER et G. MICHEL, op. cit., pp. 230-231.

2) Il est en outre intéressant de constater que le calcul concret de cette proportion passe par un dénombrement, non des substantifs, mais des verbes, puisque le remplacement d'une expression verbale par une expression nominale supprime un verbe: "Gelegentlich wird auch von nominaler bzw. verbaler Ausdrucksweise gesprochen. Dann geht es zunächst nicht um einen aussergewöhnlich hohen Anteil nominaler bzw. verbaler Elemente, sondern darum, dass ein Sachverhalt verbal, etwa in Form eines Nebensatzes, oder nominal, in Form eines Satzgliedes, ausgedrückt werden kann. (...) Wird das Satzgefüge gewählt, so sind mindestens zwei finite Verbformen (eine im Hauptsatz eine im Nebensatz) vorhanden; der Hauptsatz allein enthält nur eine finite Verbform. In dem Masse, wie der einfache Satz durch immer mehr nominale Elemente erweitert wird, verändert sich das Verhältnis immer weiter zuungunsten der finiten Verbform." (Ibid., p. 231).

ci suppose en effet une répartition uniforme des nominalisations à l'intérieur des textes, alors qu'on constate que la fréquence des nominalisations n'est pas constante à l'intérieur d'un même texte. Elle donne même l'impression d'être extrêmement "capricieuse": si on se place en un point donné du texte, le dénombrement des nominalisations dans la partie du texte qui vient d'être lue semble ne fournir aucune indication quant au nombre de nominalisations qu'on va rencontrer dans la suite: on pourra aussi bien n'en rencontrer aucune, ou une, ou plusieurs, etc.. Et si par exemple on compare des échantillons de 100 pages pris chez différents auteurs, la proportion de nominalisations variera considérablement en fonction de la place de l'échantillon dans le texte¹. Il serait donc absurde de vouloir utiliser cette méthode simple de comparaison, qui n'a de sens que si on l'accompagne d'une étude de la distribution des nominalisations dans les textes étudiés. En outre, ces comptages élémentaires mettent plutôt en question l'impression initiale: il ne semble pas que les textes de Heine contiennent beaucoup de nominalisations².

L'étude de la fréquence des nominalisations débouche donc immédiatement sur une étude statistique.

2) Ou bien la spécificité des nominalisations est qualitative, et non quantitative comme dans le cas précédent. Le lecteur est frappé, non par la masse des nominalisations, mais par leur "incongruité" ou par leur importance dans la "stratégie" du texte. L'apparition de la nominalisation peut être inattendue: le déroulement du texte laisserait prévoir en certains points l'apparition d'une expression verbale, et c'est une expression nominale qu'on rencontre. Il y a donc rupture entre un enchaînement attendu et un élément qui vient briser cet enchaînement. La pertinence stylistique des nominalisations répond dans ce cas au principe de l'"attente déçue" ou de l'écart, tel qu'a pu le formuler par exemple M. Riffaterre³. On a pu montrer que ce principe de l'attente déçue suppose qu'on admette l'existence dans le texte de chaînes probabilistes du type des chaînes de Markov⁴. Toute étu-

1) Dans la Romantische Schule de R. Haym, qui comporte environ 500 pages, ce nombre varie approximativement de 100 à 300 selon l'échantillon de 100 pages choisi.

2) Les pourcentages sont approximativement les suivants: Heine utilise environ 1 BN^v pour 100 groupes nominaux, alors que le pourcentage est de 1,4% chez Gutzkow, 1,7% chez H. Laube, 1,6% chez Hegel, etc..

3) M. RIFFATERRE, Essais de stylistique structurale. Présentation et traduction de D. DELAS. Paris 1971.

4) Cf. par exemple R. JAKOBSON, Questions de poétique. Paris 1973, pp. 485 et suiv..

de rigoureuse de cette hypothèse suppose donc qu'on recherche si l'enchaînement des nominalisations obéit à un processus probabiliste simple.

On peut enfin considérer que les nominalisations sont importantes dans la stratégie du texte: elles correspondent à une "intention de communication" précise de la part de l'auteur, et leur apparition est fortement liée au contexte. Ceci revient à admettre que les nominalisations n'apparaissent pas "au hasard", mais pour des motifs précis, aisément répertoriables. Vérifier cette hypothèse suppose qu'on puisse évaluer ce "hasard", ou l'écart entre la distribution réelle des nominalisations et la distribution que produirait "le hasard", bref qu'on étudie les nominalisations en termes statistiques ou probabilistes.

On voit que les jugements intuitifs portés sur la pertinence stylistique des nominalisations mettent en oeuvre des postulats probabilistes implicites. D'où l'idée de recourir explicitement à un outil probabiliste, et de ramener l'étude stylistique des nominalisations à l'étude de leur distribution.



Une telle étude statistique n'est possible que si l'on soumet préalablement le texte à un certain "habillage". Certes, un texte constitue par nature une séquence linéaire d'éléments qui s'enchaînent dans le temps. Toutefois, il ne saurait être question de saisir le temps "réel" de la production du texte, pas plus que celui de sa lecture, mais seulement de disposer d'une unité de référence qui soit la même pour tous les textes étudiés et qui permette un découpage homogène. Il aurait été possible de choisir une unité purement arbitraire: séquence de signes typographiques, ou de mots, ou de lignes, etc.. Mais ce procédé aurait supposé qu'on puisse donner le même support matériel à tous les textes étudiés, ce qui aurait considérablement alourdi le travail. En outre, rien ne prouvait au départ qu'une telle unité aurait la même valeur dans tous les textes étudiés.

Compte tenu de la définition qui a été donnée des nominalisations, on peut choisir comme "unité de temps" le groupe nominal, dont la définition est constante et indépendante des textes étudiés. Comme les nominalisations sont de vrais groupes nominaux, il s'agira toujours d'étudier, parmi l'ensemble des groupes nominaux, la sous-classe de ceux qui répondent à la définition de la nominalisation: soit qu'on recherche combien de groupes nominaux quelconques séparent l'apparition de deux nominalisations, soit

qu'on compte combien une séquence donnée de groupes nominaux contient de nominalisations.

En choisissant le groupe nominal comme unité minimale de temps, on réduit donc le texte à une succession de groupes nominaux en faisant abstraction de tout le reste, et on assimile l'apparition d'une nominalisation à l'intérieur des groupes nominaux du texte à l'apparition d'un phénomène dans le temps. Pour chaque texte étudié, l'"habillage" préalable consiste à compter les groupes nominaux. Ceci permet soit de découper le texte en tranches (équivalentes à une unité de temps) en comptant combien chaque tranche contient de nominalisations, soit de mesurer les écarts entre deux nominalisations. On peut alors comparer la distribution réelle des nominalisations (ou leur enchaînement) à la distribution (ou l'enchaînement) théorique correspondant à une loi statistique précise choisie comme "hypothèse nulle".

Les nominalisations sont relativement peu abondantes chez Heine; en d'autres termes, chaque texte se divisera en un grand nombre de tranches contenant chacune un petit nombre de nominalisations. D'autre part, l'enchaînement des nominalisations semble être capricieux et imprévisible: la connaissance du nombre de nominalisations apparaissant dans une portion restreinte du texte ne semble pas permettre de prévoir combien il en apparaîtra dans la portion immédiatement suivante. Ces deux données suggèrent de confronter l'enchaînement des nominalisations à un processus de Poisson. En effet, il s'agit d'un processus "sans mémoire", où la probabilité d'apparition d'un événement à un instant quelconque est indépendante du temps. Pratiquement, la procédure d'ajustement a été la suivante:

1) Pour la loi du nombre d'apparitions du phénomène par unité de temps, j'ai choisi comme unité de temps une séquence de 100 groupes nominaux. J'ai recherché le nombre moyen $\underline{m}$ de nominalisations contenues dans 100 groupes nominaux, et calculé la distribution théorique correspondant à la loi de Poisson de paramètre $\underline{m}$. Par un test du χ^2 , j'ai estimé la distance entre la distribution réelle et la distribution théorique. L'ajustement se résume donc à la valeur de $\underline{m}$ et à la valeur de la probabilité $\underline{P}$ que la distance entre la distribution théorique et la distribution réelle soit due au hasard. J'ai fixé le seuil de rejet à 5%.

Exemple d'ajustement: H. Heine, Zur Geschichte der Religion und Philosophie.
Distribution des BN^V .

Le texte comporte $n = 112$ lots de 100 groupes nominaux. Chaque lot contient $0, 1, 2, \dots, k$ BN^V . Pour les 112 lots, chaque modalité a les effectifs $n_0, n_1, \dots, n_k$. Le nombre moyen de BN^V par lot est $m = \frac{n_i \cdot i}{n}$. La fréquence

d'une modalité est $f_i = \frac{n_i}{n}$. Les fréquences théoriques correspondantes sont données par la loi de Poisson de paramètre $\underline{m}$:

$$f_i = \frac{e^{-m} \cdot m^i}{i!}.$$

On a donc le tableau suivant:

Modalités	Eff. ré- els n_i	Nbre de nom. corr.	Fréquence théor. f_i	Effectifs théor. nf_i	Distance entre eff. réel et eff. théor.
0	40	0	0,352	39,41	0,009
1	41	41	0,368	41,20	0,001
2	21	42	0,192	21,50	0,016
3	8	24	0,067	7,50	
4	1	4	0,017	1,96	
5	0	0	0,004	0,41	0,110
6	1	6	0,001	0,07	
7	0	0	0,000	0,00	
	112	117	1,001	112,05	0,136

$$\underline{m} = 1,0446$$

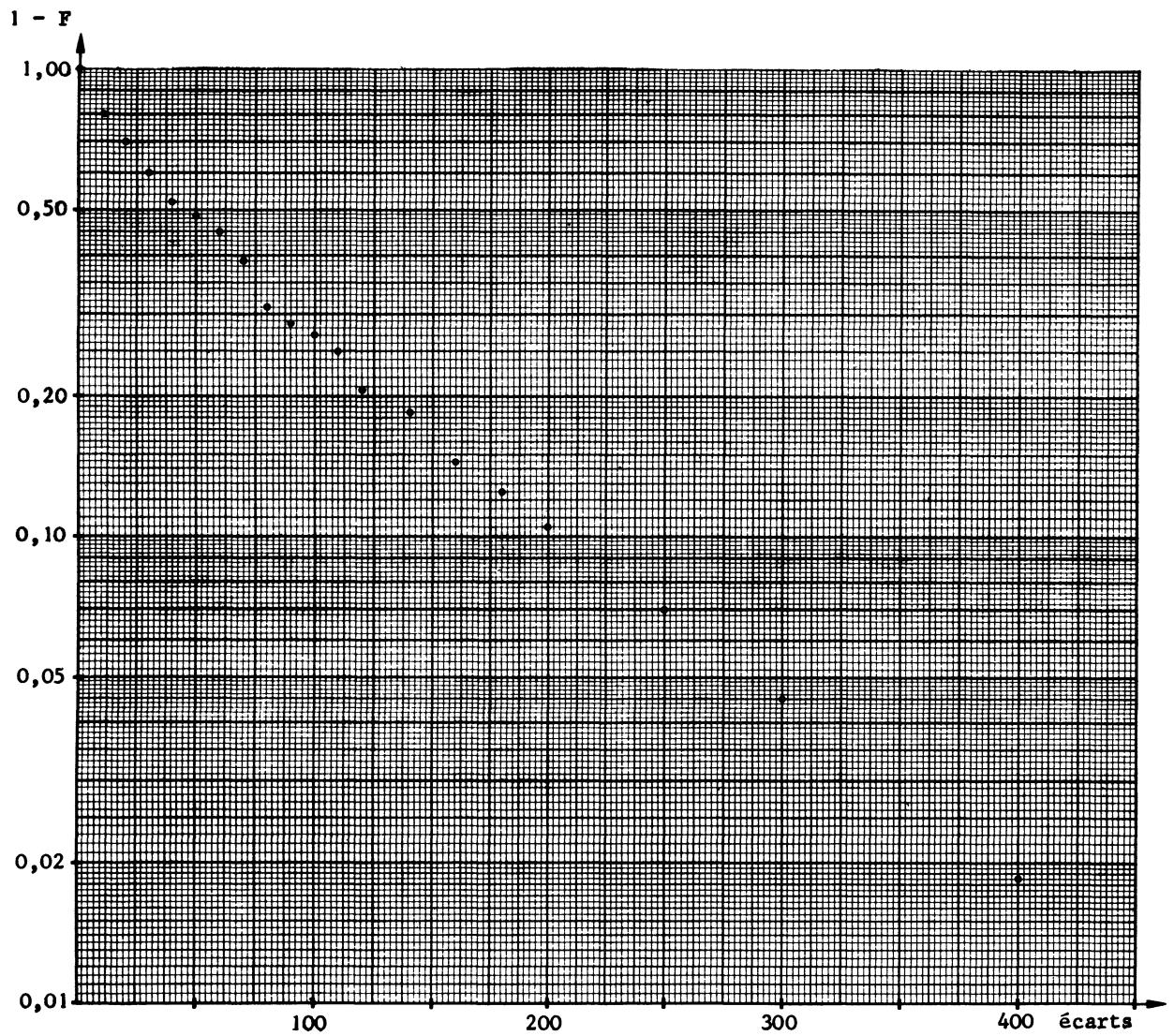
$$0,90 < P < 0,95.$$

2) Pour la loi des écarts, j'ai procédé à un ajustement graphique. Dans le cas d'un processus de Poisson, si on désigne par $\underline{x}$ les intervalles de temps séparant deux nominalisations, et par $\underline{F}$ la fréquence cumulée de ces intervalles, $\log(1 - \underline{F})$ est une fonction linéaire de $\underline{x}$. Il suffit donc de porter sur du papier semi-logarithmique les points de coordonnées $\underline{x}$ et $(1 - \underline{F})$ et de vérifier s'ils sont alignés.

Exemple d'ajustement: H. Heine, Die Romantische Schule. Distribution des écarts entre les $\underline{BN}^V$. (Figure 1).

Il est évidemment nécessaire de pouvoir comparer la distribution des nominalisations dans des textes différents, appartenant soit au même genre littéraire, soit à des genres différents. En outre, il est souhaitable de pouvoir comparer la distribution des nominalisations à celle d'autres unités linguistiques, en particulier lexicales.

L'ajustement a été effectué pour les unités suivantes: les $\underline{BN}^V$, les $\underline{GN}^Q$ s, les $\underline{GN}^Q$ das, les groupes nominaux abstraits d'origine verbale, cer-



H. Heine, Die Romantische Schule. Distribution des écarts entre les BN^V .

Figure 1

tains dérivés lexicalisés, certains lexèmes de fréquence relativement élevée dans les textes de Heine (par exemple les lexèmes en -ismus et les lexèmes en -tum).

Il a été effectué d'autre part pour les textes suivants:

H. Heine, Die Romantische Schule (RS)

Zur Geschichte der Religion und Philosophie (ZG)

Die Nordsee (N)

R. Haym, Die Romantische Schule (Haym)

R. Huch, Blütezeit der Romantik (Huch)

W. Benjamin, Ursprung des deutschen Trauerspiels (Benjamin)

Th. Mann, Der Tod in Venedig (Mann)

S. Freud, Das Unheimliche (Freud)

C.G. Jung, Bewusstes und Unbewusstes (Jung)

H. Hartig, Lärmbekämpfung in der Industrie (LI)

c'est-à-dire des textes de critique littéraire datant du XIXe siècle (H. Heine, R. Haym, R. Huch) ou contemporains (W. Benjamin), un texte littéraire (Th. Mann), des textes semi-techniques ou scientifiques (S. Freud et C.G. Jung) et un texte technique moderne (H. Hartig).

RESULTATS:

1) Distribution par unité de temps:

a) Les BN^V :

RS	m = 0,9307	$0,80 < P < 0,90$
ZG	m = 1,0446	$0,90 < P < 0,95$
N	m = 0,966	$0,50 < P < 0,70$
Haym 1 (5000 groupes nominaux)		
	m = 2,240	$P > 0,95$
Haym 2 (10000 groupes nominaux)		
	m = 2,033	$0,20 < P < 0,30$
Huch	m = 1,028	$0,50 < P < 0,70$
Mann	m = 0,704	$P > 0,95$
Freud	m = 2,906	P voisin de 0,50
Jung	m = 1,525	$0,30 < P < 0,50$
Benjamin	m = 3,680	$0,90 < P < 0,95$
LI	m = 11,733	$0,70 < P < 0,80$

b) Les GN^0 s:

RS	m = 0,554	$P < 0,001$
ZG	m = 0,667	$0,01 < P < 0,02$
N	m = 0,666	$0,30 < P < 0,50$
Haym 1	m = 1,440	$0,05 < P < 0,10$
Haym 2	m = 0,825	$P < 0,001$
Huch	m = 1,885	P voisin de 0,01
Mann	m = 0,728	P voisin de 0,01
Freud	m = 0,625	$0,90 < P < 0,95$
Jung	m = 1,000	$0,20 < P < 0,30$
Benjamin	m = 1,100	$0,01 < P < 0,02$
LI	m = 0,600	$0,01 < P < 0,02$

c) Les GN^0_{das} :

RS	m = 0,761	$P < 0,001$
ZG	m = 0,964	P voisin de 0,01
N	m = 0,966	$0,02 < P < 0,05$
Haym 1	m = 2,320	$P < 0,001$
Haym 2	m = 1,850	$P < 0,001$
Huch	m = 2,157	$0,02 < P < 0,05$
Mann	m = 1,512	$P < 0,01$
Freud	m = 5,219	$0,01 < P < 0,02$
Jung	m = 2,675	$0,30 < P < 0,50$
Benjamin	m = 1,340	$0,70 < P < 0,80$
LI	m = 0,200	P voisin de 0,50

d) Résultats annexes:

J'ai tenté ces ajustements pour étudier le rapport entre la définition linguistique des unités et leur distribution. Ont été examinés en particulier:

- l'efficacité d'une distinction, à l'intérieur des dérivés de qualitatifs, entre "abstrait" et "vrais" GN^0 ¹;
- la distribution des abstraits d'origine verbale²;

1) Cf. ici même p.40 . On peut considérer die Schönheit comme un abstrait et die ausserordentliche Dankbarkeit des Vaters für die Rettung des Kindes comme une nominalisation.

2) Par exemple das Denken, das Dichten, etc.

- l'efficacité d'une distinction, à l'intérieur des GN^Qdas, entre les GN^Qdas à valeur générique et les GN^Qdas "quelconques"¹;
- divers ajustements, portant pour la plupart sur des unités lexicales.

1) Les GN^Qs:

a) GN^Qs abstraits:

RS	m = 1,561	P < 0,01
ZG	m = 1,057	P < 0,01
N	m = 0,833	0,05 < P < 0,10
Haym 1	m = 1,566	P < 0,01
Mann	m = 2,136	P < 0,01
Freud	m = 2,344	0,10 < P < 0,20

b) GN^Qs "véritables":

RS	m = 0,438	P < 0,01
ZG	m = 0,600	0,01 < P < 0,02
N	m = 0,666	0,30 < P < 0,50
Haym 1	m = 1,245	0,02 < P < 0,05
Mann	m = 0,728	P voisin de 0,01
Freud	m = 0,625	0,90 < P < 0,95

2) Abstraites d'origine verbale comportant le suffixe das -en:

ZG	m = 0,728	0,50 < P < 0,70
N	m = 1,200	P < 0,01
Haym 1	m = 0,774	0,05 < P < 0,10
Mann	m = 1,432	0,05 < P < 0,10
Freud	m = 0,906	0,01 < P < 0,02

3) Les GN^Qdas:

a) GN^Qdas "quelconques":

ZG	m = 0,443	0,01 < P < 0,02
N	m = 0,466	0,05 < P < 0,10
Mann	m = 0,481	P < 0,001
Freud	m = 1,812	0,05 P 0,10

b) GN^Qdas à valeur générique:

RS	m = 0,715	P < 0,01
----	-----------	----------

1) On appelle "GN^Qdas à valeur générique" des groupes nominaux comme das Kindliche ou das Unheimliche, dans lesquels le défini a une valeur générique; on appelle "GN^Qdas quelconques" des groupes comme das Schöne daran, qui ne créent pas une "espèce", mais renvoient à un groupe nominal particulier. Cf. J.L. LEBRAVE, op. cit., pp. 112 à 118.

ZG	m = 0,257	$P < 0,01$
N	m = 0,500	$0,01 < P < 0,02$
Haym 1	m = 2,170	$P < 0,01$
Mann	m = 0,803	$P < 0,001$
Freud	m = 3,219	$P < 0,01$

4) Divers:

a) Groupes nominaux lexicalisés comportant le suffixe die -ung:

RS	m = 2,392	P voisin de 0,05
Freud	m = 8,187	$0,50 < P < 0,70$

b) Groupes nominaux comportant le suffixe -ismus:

RS	m = 0,692	$P < 0,01$
----	-----------	------------

c) "Adjectifs substantivés" masculins:

Mann	m = 1,805	$P < 0,01$
------	-----------	------------

d) Ensemble des "adjectifs substantivés":

RS	m = 1,077	$0,01 < P < 0,02$
Mann	m = 3,284	P voisin de 0,01

2) Loi des écarts:

Il n'est pas possible de reproduire ici l'ensemble des ajustements graphiques¹. Il suffit de savoir qu'ils apportent une bonne confirmation des ajustements numériques; on verra ci-dessous dans l'interprétation des résultats comment les unités étudiées se comportent vis-à-vis de ces ajustements.

INTERPRETATION DES RESULTATS:

1) La confrontation à un processus de Poisson permet d'établir une coupure nette à l'intérieur des unités étudiées, ceci quels que soient les textes soumis à l'ajustement statistique. On trouve en effet:

- d'une part, des populations pour lesquelles l'ajustement à une distribution poissonnienne ne peut être rejeté. Il s'agit toujours des

1) On les trouvera dans J.L. LEBRAVE, op. cit., pp. 300 à 361.

BN^V et d'eux seuls (valeurs de P élevées; bonne approximation par une droite);

- d'autre part, le reste des populations étudiées, c'est-à-dire les nominalisations de qualificatifs et les unités lexicales, pour lesquelles l'ajustement est en général difficilement acceptable (valeurs de P faibles; mauvais ajustement graphique à une droite).

C'est la double "surprise" révélée par cette étude: l'"hypothèse nulle" ne peut être rejetée dans le cas des nominalisations proprement dites (BN^V), mais d'autre part elle ne rend pas compte de la distribution des autres populations étudiées¹.

2) On peut donc considérer que la distribution des BN^V est aléatoire, et que ce trait les distingue des autres unités linguistiques étudiées². Il n'est peut-être pas inutile de rappeler ici les cas dans lesquels on recourt habituellement à la loi de Poisson: files d'attente devant un guichet de métro en dehors des heures de pointe, files d'attente de voitures à un feu rouge, arrivées de navires dans un port, etc.. Aucun voyageur n'est là par hasard, mais pour des motifs précis qu'il pourrait énumérer. Toutefois, ces raisons varient d'un voyageur à l'autre, elles sont indépendantes les unes des autres, et négligeables par rapport au phénomène pris dans son ensemble. On peut considérer qu'il en va de même pour les nominalisations: il y a autant de cas particuliers que de nominalisations particulières dans un texte; mais seule une loi aléatoire générale permet de rendre compte de l'agencement d'ensemble de ces cas particuliers.

Ceci a deux conséquences pour l'interprétation stylistique des nominalisations. En premier lieu, l'apparition d'une nominalisation en un point donné du texte est indépendante de tout ce qui a précédé. Tout se passe donc comme si l'auteur ne gardait aucun souvenir des nominalisations qu'il utilise: en particulier, il est probable qu'il ne cherche pas à les utili-

1) Au vu des premiers résultats, on aurait pu se demander si la loi de Poisson ne s'applique pas à n'importe quelle unité de fréquence faible. En effet, une population statistique arbitrairement définie - par exemple les mots commençant par f et comportant plus d'une syllabe - se prête, aussi bien que les BN^V , à un ajustement à une distribution poissonnienne. L'existence d'unités pour lesquelles l'ajustement est inacceptable est donc aussi décisive que la possibilité d'effectuer l'ajustement pour les BN^V .

2) Il serait plus exact de dire que tout se passe comme si la distribution des BN^V était aléatoire. Du moins ceci suffit-il à montrer qu'il est imprudent de postuler la pertinence immédiate de cette distribution, comme le font les études stylistiques intuitives.

ser systématiquement, ni à en limiter systématiquement l'emploi¹. Toute intervention systématique aurait en effet pour résultat d'accroître la distance entre la distribution réelle et la distribution théorique des nominalisations.

En second lieu, tout se passe comme si la distribution des nominalisations était indépendante du contexte: s'il en était autrement, la spécificité des contenus et des intentions de communication viendrait spécifier la distribution des nominalisations.

Il est donc illusoire de rechercher une interprétation thématique ou stylistique pour la distribution des BN^V dans une oeuvre donnée. En revanche, il serait certainement intéressant de comparer un grand nombre d'auteurs du point de vue de la fréquence des nominalisations. Si on admet que la loi de Poisson rend compte de la distribution des nominalisations quelle que soit l'oeuvre et quel que soit l'auteur, l'emploi des nominalisations est défini exhaustivement par le paramètre de la loi de Poisson correspondante. Dans ces conditions, on peut chercher quelle est la relation entre les caractéristiques d'une oeuvre (sujet, période, auteur) et la valeur prise par ce paramètre. Il est par exemple intéressant de voir que la fréquence des nominalisations est beaucoup plus élevée dans les textes techniques ou scientifiques que dans les textes littéraires: Freud utilise beaucoup plus de nominalisations que Th. Mann. De même, on constate que les BN^V sont beaucoup plus fréquents dans l'allemand contemporain que dans l'allemand du XIXe siècle.

3) Inversement, on peut rejeter l'hypothèse d'une distribution aléatoire pour les 'nominalisations de qualitatifs', dont la répartition est beaucoup plus proche de celle des unités lexicales. Toutefois, il serait prématuré d'en tirer des conclusions stylistiques: on a vu que la définition linguistique des ' GN^Q s' est beaucoup plus ambiguë que celle des BN^V , puisqu'il est souvent difficile de dire s'il s'agit d'une nominalisation ou d'un abstrait qualitatif². Si l'on récusait l'existence d'une correspondance entre ces unités et des expressions verbales, les GN^Q s ne seraient

1) Ceci semble confirmé par l'étude des variantes dans les manuscrits de Heine: alors qu'on y voit sans cesse à l'oeuvre le principe stylistique qui commande d'éviter les répétitions, les nominalisations ne semblent l'objet d'aucune attention spécifique. Il ne semble pas qu'il y ait un jeu de variantes caractéristique des nominalisations.

2) die Schönheit des Gedichts peut être paraphrasé de deux façons différentes: das Gedicht ist schön, mais aussi das Gedicht "hat" Schönheit.

qu'un type particulier d'unités lexicales. Quant aux GN^Q_{das} , il est clair que la correspondance qu'ils entretiennent avec les groupes verbaux du type was + an + groupe nominal + groupe qualitatif + ist n'est pas de même nature que celle qui caractérise les BN^V ¹.

° °

°

Ici s'arrête forcément l'interprétation des résultats. En effet, une interprétation plus poussée n'est possible que si l'on connaît d'autres unités linguistiques dont la distribution est du même type que celle des nominalisations. On pourrait envisager une coupure à l'intérieur de la mise en forme linguistique des énoncés: étant donné une "intention de communication" particulière (portant par exemple sur la fondation de Rome par Romulus), l'auteur peut choisir entre différents énoncés équivalents du point de vue référentiel: Romulus a fondé Rome, Rome a été fondée par Romulus, Rome a eu Romulus pour fondateur, la fondation de Rome par Romulus, etc., mais différents par la mise en forme ². La généralisation d'une étude comme celle qui a été effectuée sur les nominalisations permettrait de sélectionner dans la mise en forme linguistique les opérations qui sont l'objet d'un choix - conscient ou non - de la part de l'auteur. Si on admet que les unités lexicales, à la différence des nominalisations (BN^V), n'ont pas une distribution aléatoire, on pourrait en conclure que les choix de l'auteur (sur le plan du contenu, mais aussi sur celui du "style") interviennent au niveau de la substance du contenu, mais non à celui de sa forme.

Mais rien n'atteste que c'est bien dans ce cadre qu'il faut interpréter le caractère aléatoire de la distribution des nominalisations. On pourrait aussi chercher un lien entre ce type de phénomènes et les faits d'énonciation. En effet, on peut définir les nominalisations comme des énoncés dont on a retranché les marques d'énonciation propres au groupe

1) Il est peut-être intéressant de signaler que l'ajustement statistique a servi de révélateur pour cette diversité linguistique des unités étudiées. Le choix pour les BN^V d'une définition en termes de "complexes de signifiés" a été suggérée en partie par l'ajustement statistique, qui a permis de découvrir les inconséquences linguistiques commises lors de l'étiquetage des nominalisations. Il était facile dès lors de constater qu'une partie des groupes nominaux classés parmi les GN^S comportait le même type d'anomalie.

2) Cf. par exemple J. FOURQUET, Qu'est-ce que la linguistique? In: Journal de Psychologie normale et pathologique. Janvier-Juin 1973, pp. 15-16.

verbal. Dans cette perspective, l'auteur qui emploie une nominalisation cherche à s'effacer en tant que sujet d'énonciation. Mais cet effacement n'est interprétable que si on le rattache à l'ensemble des faits d'énonciation présents dans le texte. L'étude de la distribution des nominalisations telle qu'elle a été effectuée ici semble montrer qu'elles ne sont pas l'objet d'un choix particulier en tant que sous-classe de groupes nominaux; peut-être faudrait-il chercher si elles n'ont pas une pertinence spécifique en tant qu'énoncés sans marques d'énonciation.

° ° °