

Un index sélectif d'un texte philosophique

HUSSERL, *Die Idee der Phänomenologie*

Étienne ÉVRARD et Gérald PURNELLE

Abstract

To assist the content analysis of a philosophical text, we have created a method of constituting both a selective index reduced to philosophically significant words and several frequency lists. We have applied this method to *Die Idee der Phänomenologie* where Husserl explains for the first time his theory of the phenomenological reduction and of the constitution of the objects in the conscience. The results proved to be valuable for the philosophical interpretation.

Edmund Husserl (1859–1938) est sans contredit l'un des plus grands noms de la philosophie contemporaine, et le courant phénoménologique, dont il est le père, a marqué et marque encore puissamment l'activité philosophique du XX^e siècle¹.

Mais deux raisons rendent difficile l'accès de sa pensée. Tout d'abord, Husserl a souvent eu de longues hésitations avant de publier le fruit de ses réflexions. Après quelques premiers écrits et les *Logische Untersuchungen* (1900–1901), où apparaissent les premiers éléments de la phénoménologie, il n'édite rien, sinon, en 1911, un article important mais assez bref, jusqu'à la parution, en 1913, du 1^{er} livre des *Ideen zu einer reinen Phänomenologie*. Les 2^e et 3^e livres de cet ouvrage ne seront jamais publiés par Husserl, bien qu'il y ait beaucoup travaillé. De semblables hésitations se répètent par la suite. Rappelons seulement le cas des *Méditations cartésiennes* : invité à Paris, Husserl y fit, en 1929, des conférences très remarquées qu'il prononça en allemand; les organisateurs lui

¹ On trouvera une esquisse rapide de la carrière de Husserl par exemple dans Arion L. KBLKBL et René SCHERER, *Husserl*, coll. «Sup-Philosophes», Paris, P.U.F., 1971.

✉ Université de Liège; CIPL-LASLA; place du 20-Août, 32; B-4000 Liège (Belgique).
Fax : + 32 41 23 25 45 et + 32 41 66 57 02 E-mail : u0013a1@bliulg11

MOTS-CLÉS : Analyse du contenu, index sélectif, liste de fréquence.

en demandèrent le texte en vue d'une traduction française qui parut en 1931; quant au texte allemand, Husserl ne le trouva jamais assez mûr pour être publié². Ces conditions ne facilitent pas la connaissance de la pensée du maître, sauf pour ses élèves et ses familiers.

Une deuxième raison complique encore la tâche. Husserl, qui était d'origine juive, tombait sous le coup de la législation instaurée par le régime nazi. En raison de sa notoriété internationale, sa radiation de l'Université, après avoir été annoncée, fut promptement rapportée. Mais il resta soumis à des contrôles. Ses seules manifestations publiques (conférences, articles) eurent désormais lieu hors d'Allemagne. La situation devint telle que, pour permettre la conservation et l'utilisation de son *Nachlass*, il en établit avec ses assistants un classement systématique matérialisé par des sigles et pensa à le mettre à l'abri en le faisant transporter à l'étranger³.

Ce projet ne se réalisa qu'après sa mort. En septembre 1939, le franciscain belge H.L. Van Breda, de l'Université de Louvain, en accord avec la veuve et avec le dernier assistant de Husserl, réussit à transférer clandestinement les notes, les manuscrits et la bibliothèque du philosophe. Au mois d'octobre suivant, la Fondation Franqui accorda les moyens financiers nécessaires à la création et au fonctionnement des Archives Husserl⁴. Immédiatement commença la transcription des notes et manuscrits (écrits pour la plupart en sténographie) par les deux derniers assistants de Husserl. Mais l'occupation de la Belgique entraîna l'éloignement de ces derniers et obligea à camoufler les Archives. La transcription continua cependant, mais clandestine et lente. Après la guerre, le travail reprit à un rythme normal et le projet s'élabora de procurer des éditions critiques non seulement des inédits mais aussi des textes édités du vivant de l'auteur, avec les corrections qu'il avait notées sur ses exemplaires et, en ajout, des textes inédits relatifs aux mêmes sujets. Les deux premiers volumes de cette collection, appelée *Husserliana*, parurent en 1950⁵.

Inutile de dire que cette série de publications (il y a une trentaine de volumes à l'heure actuelle) ont suscité un regain d'intérêt pour Husserl tout en procurant une meilleure connaissance de ses idées.

² Sur les publications de Husserl et sur ses inédits, v. l'ouvrage cité n. 1, pp. 85 sq. Voir aussi Sabine MÖDERSHEIM, *L'exploitation de l'œuvre posthume de Husserl*, dans Éliane ESCOUBAS et Marc RICHIR (éd.), *Husserl*, Grenoble, Jérôme Millon, 1989, pp. 185–197.

³ *Ibid.*, p. 186.

⁴ *Ibid.*, pp. 186 sq.

⁵ *Ibid.*, pp. 187 sq. Sur les problèmes de la transcription, v. par exemple Walter BIEMEL, *La montagne magique phénoménologique*, dans *Husserl* (cité n. 2), pp. 214–218.

C'est le deuxième volume des *Husserliana* que nous avons pris pour objet de la présente expérience. Sous le titre *Die Idee der Phänomenologie*, on y trouve cinq leçons que Husserl professa à Göttingen comme introduction à un cours du semestre d'été de 1907 (du 26 avril au 2 mai)⁶. Leur importance vient de ce que le philosophe y expose pour la première fois ses théories de la réduction phénoménologique et de la constitution des objets dans la conscience, mettant ainsi un terme à l'inquiétude intellectuelle qu'il ressentait relativement à la nécessité de résoudre le problème de la critique de la raison — inquiétude qui, dit-il, l'empêchait de vivre⁷. Désormais, sa philosophie peut s'appeler une phénoménologie transcendantale⁸. C'est donc un moment capital dans l'évolution de la pensée husserlienne. De là l'importance de ce texte.

*
* *

L'analyse du contenu d'un texte philosophique, si l'on veut éviter les risques de dérives interprétatives fondées sur des impressions trop peu contrôlées, a intérêt à disposer de relevés exhaustifs du vocabulaire, ou du moins des vocables exprimant les concepts, les thèmes, les motifs qui en font la matière. Husserl, comme de nombreux philosophes, a une langue peu imagée; il en résulte que le risque d'ambiguïté y est relativement réduit : la présence ou l'absence d'un vocable y manifeste presque nécessairement celle du concept correspondant et l'importance d'un concept ou d'un thème y est en relation plus ou moins étroite avec le nombre d'occurrences du ou des vocables qui les expriment.

On est ainsi conduit au projet d'un index sélectif, réduit aux vocables philosophiquement significatifs, avec référence et dénombrement des occurrences, avec aussi classement des vocables en fréquence décroissante. Pour la réalisation d'un tel projet, on songe tout de suite à l'informatique.

Dans une entreprise de ce genre, on rencontre immédiatement un premier obstacle qui suffit à décourager bien des bonnes volontés, nous voulons dire la saisie du texte, sa mise en mémoire électronique. S'il faut introduire le texte à partir du clavier d'un terminal ou d'une PC, il est à craindre que le chercheur soit rebuté et arrête tout de suite les frais. Pour notre part, nous avons opté pour la lecture optique (*Optical Character Recognition*) au moyen d'un scanner. La

⁶ Edm. HUSSERL, *Die Idee der Phänomenologie, Fünf Vorlesungen*, éd. et introd. de Walter BIEMEL, Haag, Martinus Nijhoff, 1950 (*Husserliana*, II).

⁷ Voir le texte allégué dans l'ouvrage cité à la note précédente, pp. VII sq.

⁸ W. BIEMEL, introd. de l'éd. citée n. 6, pp. VII-IX.

typographie des *Husserliana* est heureusement assez claire, nette et homogène pour se prêter à cette procédure. Les caractères ou groupes de caractères que le logiciel de lecture optique s'est révélé incapable d'identifier seul furent relativement peu nombreux. Une fois la lecture achevée, il fallut cependant faire la toilette du texte en utilisant chaque fois que c'était possible des moyens automatiques : réunir les deux éléments des mots répartis sur deux lignes en les plaçant entièrement sur la seconde (ceci pour permettre une référencement correcte); introduire, avec des codes distinctifs, les numéros des leçons ainsi que la pagination (les fins de lignes, elles, avaient été automatiquement marquées lors de la lecture optique); resserrer les mots imprimés en espacés (pour éviter que chaque lettre ne soit considérée ultérieurement comme un mot); éliminer les tentatives faites par le logiciel pour interpréter les annotations marginales de l'exemplaire utilisé; rechercher automatiquement les caractères ou groupes de caractères pour lesquels il y a des risques de lecture fautive et corriger éventuellement au cas par cas : assez souvent, le logiciel prend la lettre ℓ pour un I ; par ailleurs, il interprète tant bien que mal le *scharfes s* (β), qu'il ne connaît pas et que, pour notre part, nous représentons par $\$$; nous avons donc recherché automatiquement ces transcriptions approximatives et, après vérification sur le texte même, les avons éventuellement corrigées. Ceci fait, un listage du texte en mémoire a servi à une relecture attentive destinée à éliminer le plus complètement possible les erreurs restantes. Toutes ces opérations se réalisent en un temps assez court : pour *Die Idee*, qui compte 50 pages (en ne comptant que les pages du texte proprement dit, c'est-à-dire en négligeant l'introduction, les pages de titre et les annexes), cela ne nous a guère demandé que trois petites après-midi.

Une fois le fichier-texte purgé de ses fautes, un programme approprié en tire automatiquement un deuxième fichier, dans lequel chaque occurrence fait l'objet d'un *record* divisé en quatre zones, la première laissée temporairement vide, la deuxième contenant la forme, d'abord sous son aspect normal puis réduite à ses composants alphabétiques (c'est-à-dire sans les éléments secondaires tels les accents, les trémas, la distinction majuscule/minuscule) notés en capitales, la troisième réservée à la référence et comportant trois zones, une pour le numéro de la leçon, une pour le numéro de la page, une pour le numéro de la ligne (calculé à partir des codes de fin de paragraphe placés, lors de la lecture optique, à la fin de chaque ligne), la quatrième, enfin, recevant le numéro d'ordre de l'occurrence dans l'œuvre, calculé au fur et à mesure que le fichier est constitué. Les 14 621 mots du texte ont subi ce traitement en environ 8 minutes.

Ce fichier-occurrences est trié en ordre alphabétique des occurrences. Après cette opération, il présente une succession de *records* qui soit sont les uniques représentants de la forme qu'ils concernent, soit constituent des séries plus ou

moins longues d'occurrences ayant en commun une même forme; les séries qui sont identiques par leur composition alphabétique mais diffèrent par un ou des éléments accessoires se trouvent évidemment en succession immédiate. Le fichier ainsi trié est soumis à un programme d'impression *ad hoc* qui en tire un listage où chaque forme, au moment où elle se présente pour la première fois, est imprimée; à sa suite viennent un point-et-virgule et les références des occurrences de cette forme, imprimées en ligne; au moment où l'ordinateur, lisant le fichier, repère une forme différente, il marque la fin de la rubrique consacrée à la forme en cours en la réimprimant à la marge et en la faisant suivre d'un chiffre exprimant le nombre de ses occurrences. C'est seulement après ces opérations que la forme nouvelle, qui avait déclenché ce processus, est imprimée et que le cycle recommence. Notons encore qu'à la fin de chaque série de formes alphabétiquement identiques mais présentant des différences accessoires (par exemple *Aber* et *aber*, ou, dans l'échantillon ci-dessous, *Könnte* et *könnte*), la forme alphabétique commune s'imprime en capitales, suivie de la fréquence globale des formes ainsi regroupées. On trouvera ci-dessous un extrait de ce résultat imprimé.

können; I,24, 3; II,30,21; II,31,30; II,31,31; II,31,36; III,43, 9;
 III,45,38; III,46,17; III,49, 6; III,49,14; III,49,18; III,51,18;
 IV,56, 7; IV,56,30; IV,57,13; IV,60,20; IV,60,24; IV,61,28;
 IV,61,29; IV,62,10; IV,63,15; V,73,24

können 22

Könnte; IV,57, 9

Könnte 1

könnte; I,21,13; II,36, 9; II,37,19; II,37,31; III,46,35; III,47, 7;
 IV,60,21; V,67,12

könnte 8

KONNTE 9

könnten; IV,57, 6; IV,61,33; V,70,14

könnten 3

Konsequenterweise; II,38, 1

Konsequenterweise 1

konstatieren; III,50,11

konstatieren 1

konstituieren; III,52, 6; IV,62,15; V,68,20; V,70,35; V,71, 9; V,71, 9;
 V,71,36

konstituieren 7

konstituierend; V,67,19

konstituierend 1

konstituierende; V,75,14

konstituierende 1

konstituierenden; IV,58,11; V,70, 5

konstituierenden 2

konstituiert; II,31,22; IV,56,35; IV,58, 9; V,68, 5; V,68, 7; V,68,35;
V,71,18; V,71,20; V,71,23; V,72, 4; V,72, 6; V,75, 3; V,75,20

konstituiert 13

konstituierten; V,73,39

konstituierten 1

Konstitution; II,31,37; V,74,10

Konstitution 2

Ce que l'on a ainsi produit est un index de formes non lemmatisées ni désambiguées, avec mention des fréquences. Nous devons ici un mot d'explication à nos lecteurs. Nous avons souvent exprimé notre préférence pour les index lemmatisés; par ailleurs, nous avons à de multiples reprises critiqué l'indication des effectifs de formes non désambiguées (dans *Die Idee*, par exemple la fréquence 59 de *SEIN* concerne indistinctement un ensemble de formes dont les unes appartiennent au possessif, tandis que les autres sont des infinitifs). Notre opinion n'a pas varié sur ces deux points. Ce qui, à nos yeux, justifie cependant la production d'un index de formes avec indication de fréquence, c'est que, dans l'ensemble du traitement que nous proposons, cet index n'est qu'une étape, un instrument de travail tout provisoire. Il permet tout d'abord de vérifier une nouvelle fois la correction de la saisie : les anomalies de l'ordre alphabétique révèlent en effet les fausses interprétations orthographiques du logiciel. Ainsi, dans la première version de l'index provisoire de *Die Idee*, une occurrence de *Relativismus* avait été traitée à part et se trouvait entre *regeln* et *reicht*; c'est que le *l* y avait été pris pour un *I* par le logiciel d'OCR. Par ailleurs, le chercheur pressé, en attendant la production de l'index sélectif lemmatisé, peut déjà trouver dans l'index-formes des indications très révélatrices : le plus souvent, les diverses formes d'un même lemme sont assez proches dans l'ordre alphabétique; les cas où cela ne se vérifie pas peuvent assez facilement être repérés, si bien que le regroupement de toutes les formes d'un même lemme est une opération, fastidieuse sans doute mais relativement aisée, qui fournit une indication sur la fréquence globale du lemme; quant aux ambiguïtés qui réunissent sous une même

rubrique des formes identiques mais relevant de lemmes différents, un familier du texte considéré, à condition d'y être attentif, les repérera assez facilement. On voit donc que l'index-formes, s'il ne peut être tenu pour un produit définitivement satisfaisant, est cependant utile à condition que l'on ne perde pas de vue son caractère provisoire.

L'étape suivante conduit sur la voie de l'index sélectif lemmatisé. Elle consiste à choisir les lemmes que l'on désire sélectionner et à lemmatiser les formes correspondantes; elle consiste aussi à marquer les formes que l'on exclut de la lemmatisation. On se souvient que les *records* des occurrences commencent par une zone laissée vide jusqu'ici. Sur le *record* qui vient en tête d'une série de *records* à forme identique, on introduit dans cette zone soit le lemme, s'il s'agit d'un lemme que l'on souhaite sélectionner, soit un simple tiret (-) dans le cas contraire. Dans les *records* suivants, la zone lemme est laissée vide jusqu'au moment où apparaît une forme différente. Toutefois, pour les formes ambiguës dont l'une au moins des lemmatisations possibles doit être sélectionnée, la procédure sera plus compliquée : soit la forme *meine* qui peut relever du lemme *meinen*, à sélectionner, ou du possessif *mein*, à négliger; l'index des formes indique d'abord la référence I,20,17, qui, vérification faite sur le texte, concerne le possessif et doit être marquée du tiret; la référence suivante a trait à la forme verbale, elle reçoit donc, dans sa première zone, le lemme *meinen*; viennent ensuite trois références de possessif : seule la première est marquée du tiret; enfin, les deux dernières concernent le verbe : la première des deux reçoit le lemme.

L'opération terminée, un petit programme balaie le fichier et remplit la zone lemme des *records* successifs, à condition qu'il la trouve vide, en y dupliquant le lemme du *record* précédent. Au terme de cette opération, les zones lemme de tous les *records* sont pourvues soit d'un lemme, soit d'un tiret.

Pour obtenir l'index sélectif qui est l'objet de cet article, il convient de soumettre le fichier-occurrences à un tri complexe qui doit avoir pour effet de mettre les occurrences dans l'ordre alphabétique des lemmes et, à l'intérieur de chaque lemme, dans l'ordre du texte. En raison du code du tiret, les *records* qui ont ce caractère dans leur première zone se situent en tête. Ils concernent les occurrences relevant de lemmes non sélectionnés; on ne peut cependant les négliger purement et simplement; en vue de leur traitement ultérieur, il convient d'en former un fichier distinct. Quant à la partie lemmatisée, elle est soumise à un programme d'impression qui imprime les lemmes lors de leur première apparition et les fait suivre des références qui s'y rapportent, dans l'ordre du texte, puis d'une indication de fréquence. L'introduction d'un critère supplémentaire de tri, portant sur la forme, permettrait l'impression d'un index où chaque article réserverait une partie distincte à chacune des formes sous lesquelles se rencontre

le lemme : les références se répartiraient donc en autant de séries (en ordre du texte) qu'il y a de formes représentées dans le texte. C'est là une disposition possible, mais qui n'est guère favorable à l'analyse du contenu. C'est pourquoi nous l'avons négligée, tout en réservant la possibilité de la pratiquer.

Revenons au fichier des *records* portant un tiret en première zone. Il permet de préciser, pour l'utilisateur, les limites de la sélection opérée. Pour ce faire, il est trié en ordre alphabétique sur la forme, puis un programme d'impression imprime chaque forme à sa première apparition. En consultant la liste ainsi obtenue, l'utilisateur pourra vérifier les choix auxquels a obéi la sélection. En particulier, il pourra sans peine vérifier si le mot sur lequel il s'interroge est présent dans le texte mais a été négligé ou s'il en est complètement absent; dans cette vérification, il devra cependant être attentif au fait que le lemme auquel il s'intéresse peut se présenter sous diverses formes.

L'ensemble formé par la liste dont il vient d'être question et par l'impression des lemmes et de leurs références constitue un index sélectif parfaitement explicite concernant le contenu du texte considéré. La liste des exclusions nous paraît, de ce point de vue, revêtir une grande importance : elle seule permet une utilisation correcte des données. Par ailleurs, il convient de dire que la sélection n'a rien de définitif et que, si le besoin s'en fait sentir, il est à tout moment possible de sélectionner les formes d'un lemme qui avait d'abord été rejeté.

Indépendamment de l'impression précédemment décrite, le fichier des lemmes sert à constituer un fichier de fréquence. Ce fichier enregistre chaque lemme et le fait suivre de six chiffres concernant la fréquence dans chacune des cinq leçons et la fréquence totale (ces fréquences obtenues à partir d'un comptage automatique fondé sur le premier élément de la zone référence). Le fichier-fréquence est d'abord imprimé dans l'ordre où il se trouve, c'est-à-dire l'ordre alphabétique. Il est ensuite trié dans l'ordre de fréquence décroissante sur la fréquence dans chacune des cinq leçons et dans la totalité. Chacun de ces tris est suivi d'une impression. On dispose ainsi d'un moyen d'évaluer l'importance des vocables — et des concepts ou notions dont ils sont porteurs — tant dans l'ensemble de l'œuvre que dans chacune de ses parties et d'observer la manière dont évolue l'emploi d'un vocable donné au cours du texte. À titre d'exemples, signalons seulement que *Reduktion*, le plus souvent accompagné de l'adjectif *phänomenologische*, fait une première apparition dans la deuxième leçon, qu'il est fréquent dans la troisième (8), qu'il l'est un peu moins dans la quatrième (7), et que, dans la cinquième, il est rare (2), à quoi il convient d'ajouter les diverses formes du verbe *reduzieren*, qui se rencontrent dans la 1^{re} leçon (1), dans la 3^e (3), dans la 4^e (1) et dans la 5^e (2), ce qui donne les effectifs globaux 1, 1, 11, 8, 4. Quant aux diverses formes du verbe *konstituieren*, on n'en rencontre aucune

dans la première leçon, elles n'apparaissent presque pas dans la seconde (2) et dans la troisième (1), mais sont un peu plus fréquentes dans la quatrième (4) et deviennent très fréquentes dans la cinquième (20). On voit ainsi la place, dans l'économie générale du texte, des deux notions dont l'apparition donne à *Die Idee* une bonne part de sa nouveauté. Par ailleurs, il est intéressant d'observer que les vocables *Ausschaltung* et *ausschalten*, importants pour la théorie de la réduction, ne se trouvent ici que deux fois. En revanche, *Intention* et *intentional* sous ses diverses formes totalisent 7 occurrences et marquent ainsi la place que, dès cette époque, l'intentionnalité occupe dans la pensée husserlienne. On mesurera par contre ce que le système a d'encore inachevé si l'on observe que ni *noetisch*, ni *noematisch* n'apparaissent, non plus que *Intersubjektivität* : il s'agit là, en effet, de termes correspondant à quelques thèmes destinés à jouer un rôle considérable dans le développement de la phénoménologie. Ces quelques exemples suffisent à faire deviner tout ce que l'index peut apporter au lecteur et au critique.

Il nous reste à signaler que la méthode appliquée ici à Husserl est applicable à tout autre texte, dans quelque langue que ce soit, dès le moment où l'intérêt se porte essentiellement sur l'analyse du contenu⁹. En particulier, le nombre de parties à considérer dans le texte pour le dénombrement des fréquences peut être réduit ou augmenté sans problème. Il faut seulement dire que, pour des œuvres nettement plus longues que *Die Idee*, il serait indispensable de transférer tout le logiciel sur gros ordinateur, alors que, jusqu'ici, nous l'avons pratiqué sur PC. Un tel transfert n'a rien de très difficile.

⁹ En fait, avant de travailler sur *Die Idee*, nous avons esquissé la méthode ici décrite pour répondre à la demande d'une étudiante, Mlle Anne Staquet, qui voulait examiner la thématique d'un texte de Stig Dagerman; les enseignements puisés dans l'index-formes qui lui a été fourni l'avaient conduite à une interprétation très convaincante de ce texte, que nous espérons publier dans le prochain numéro de *Revue. ISSH*.