La théorie des surfaces algébriques dans les *Mathematische Annalen* à l'épreuve de la textométrie (1869-1898)

François Lê [*]

Postprint. Août 2021

La textométrie, parfois aussi appelée lexicométrie ou analyse de données textuelles, se présente comme une méthode d'analyse de corpus de textes portant son attention sur le lexique et les catégories grammaticales qui y sont utilisés, et alliant avec un traitement quantitatif de telles données une lecture qualitative toujours re-située des textes étudiés. Comme l'expliquent Bénédicte Pincemin et Serge Heiden, « la textométrie s'est essentiellement développée en France dans la lignée des recherches pionnières de Pierre Guiraud (1954, 1960) et de Charles Muller (1968, 1977) en statistique lexicale (évaluation de la richesse du vocabulaire d'un texte, vocabulaire caractéristique d'un texte). Elle reprend et poursuit également les méthodes d'analyse des données (analyses factorielles, classifications) mises au point par Jean-Paul Benzécri (1973) et déjà appliquées par lui aux données linguistiques [...]. La textométrie développe en outre de nouveaux modèles statistiques pour rendre compte de caractéristiques significatives de données textuelles : attirances contextuelles des mots [...], linéarité et organisation interne du texte [...], contrastes intertextuels [...], indicateurs d'évolution lexicale. » [Pincemin et Heiden 2008]. Il s'agit donc d'exploiter, avec le soutien de l'outil informatique, diverses mesures statistiques afin de déceler dans la masse des mots d'un corpus de telles caractéristiques, qu'il importe ensuite d'analyser et d'interpréter plus finement.

Ces méthodes sont aujourd'hui utilisées par nombre de chercheurs et de chercheuses dans différents champs des sciences humaines et sociales, dont l'histoire littéraire, les sciences politiques et la sociologie [1].

[*] Univ Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5208, Institut Camille Jordan, 43 blvd. du 11 novembre 1918, F-69622 Villeurbanne Cedex, France.

1. Voir par exemple les références données dans les écrits de synthèse [Lemercier

En revanche, l'histoire des sciences ne semble pas encore s'être emparée de l'outil textométrique, à quelques exceptions près. Ainsi, à l'aide d'un logiciel développé à l'occasion, Pierre Teissier, Matthieu Quantin et Benjamin Hervy [2018] ont proposé une analyse d'un corpus de transcriptions d'entretiens oraux avec des scientifiques ayant participé aux recherches sur les matériaux au cours de la deuxième moitié du XX^e^ siècle, analyse basée notamment sur une structuration du corpus en *clusters* reflétant des liens de proximité lexicale entre textes. Du côté de l'histoire des mathématiques, deux thèses récentes se sont appuyées sur des techniques textométriques pour répondre à des questions ponctuelles : repérer, à l'aide de certains termes caractéristiques prédéfinis, des recensions d'articles qui relèveraient de la géométrie, [Duran 2019, p. 122], ou identifier des thématiques de textes sur la base des noms et adjectifs employés dans leurs titres, [Cléry 2020, p. 179 sqq.].

Par sa nature même, la textométrie renvoie à deux facettes des recherches en histoire des mathématiques — l'attention portée sur le vocabulaire et les méthodes quantitatives — qui semblent avoir été jusqu'à présent thématisées à des degrés différents. Ainsi, si le lexique d'un corpus dans son ensemble n'a jamais été l'objet propre des investigations, toutes celles-ci ne portent pas moins un intérêt indéniables aux mots, qui servent toujours d'une manière ou d'une autre — et de façon plus ou moins explicite — à détecter et étudier divers phénomènes, à constituer des corpus pertinents ou à catégoriser des textes, par exemple [2]. Au contraire, les méthodes quantitatives sont plus souvent expressément introduites et discutées en tant que telles ; depuis les années 1990, elles sont régulièrement mobilisées en histoire des mathématiques, que ce soit pour étudier des sociétés savantes, des journaux, des populations ou des

et Zalc 2008, ch. IV ; Brunet 2016 ; Lebart, Pincemin et Poudat 2019, ch. I], ou les actes des Journées internationales d'analyse statistique de données textuelles, disponibles sur http://lexicometrica.univ-paris3.fr. Des recherches sur la question du lexique (actuel) des sciences, surtout humaines et sociales, sont par ailleurs regroupées dans [Tutin 2007 ; Jacques et Tutin 2018]. Enfin, notons que des explications et des réponses aux plus courantes des nombreuses objections qui ont été (et sont encore) soulevées à l'égard de la textométrie sont apportées avec efficacité dans [Pincemin 2020].

2. En histoire des sciences, des recherches collectives menées dans les années 1980 se sont intéressées au lexique scientifique et se sont vues concrétisées dans les neuf *Documents pour l'histoire du vocabulaire scientifique*, publiés entre 1980 et 1989. Y sont incluses diverses études de mots ou de groupes de mots au sein de corpus donnés, de glossaires et d'autres lexiques techniques, ou encore de processus de formation et d'évolution de lexiques spécialisés. Le traitement de ces études, cependant, n'est pas de nature textométrique. Retenons également que le lexique du *Novum organum* de Francis Bacon a fait l'objet de recherches plus proches de la voie textométrique, présentées dans [Fattori 1980].

domaines mathématiques particuliers [3].

C'est dans cette dernière perspective que se situe le présent article, qui propose d'aborder une partie de la géométrie algébrique du dernier tiers du XIX^e siècle à l'aide d'outils textométriques.

Comme on le sait, certains des travaux quantitatifs évoqués à l'instant ont souligné l'importance de se munir de divers garde-fous méthodologiques dans une telle entreprise, par exemple pour repérer ce qui relèverait de la géométrie algébrique à l'époque considérée. Or, l'étiquette « géométrie algébrique » elle-même se trouve rarement employée au XIX^e siècle, et n'apparaît en particulier ni dans les diverses divisions disciplinaires du *Jahrbuch über die Fortschritte der Mathematik*, ni dans l'index du *Répertoire bibliographique des sciences mathématiques*, ni même dans le *Catalogue of scientific papers*. On la trouve cependant dans le *Generalregister* des *Mathematische Annalen*, registre paru en 1898 en même temps que le cinquantième volume de la revue et proposant entre autres une classification disciplinaire rétrospective des articles publiés depuis le premier volume de 1869 [4].

La géométrie algébrique est une des trois sections du chapitre consacré à la géométrie, intercalée entre les « disciplines purement géométriques » et la « géométrie différentielle ». Les titres de ses dix paragraphes en font voir les différents thèmes constituants : courbes coniques et d'ordre supérieur, surfaces quadriques et d'ordre supérieur, correspondances projectives et transformations algébriques en lien avec les configurations, géométrie d'éléments supérieurs de l'espace et « géométrie algébrique » d'espaces de dimensions supérieures.

J'ai choisi de m'intéresser ici aux textes classifiés dans les deux paragraphes renvoyant explicitement aux surfaces algébriques d'ordre supérieur, l'un étant consacré aux « surfaces algébriques spéciales », l'autre à la « théorie générale des surfaces algébriques supérieures [5] ». De

3. La liste de ces travaux étant désormais assez longue et plutôt bien connue, je me contenterai ici de renvoyer au bilan fait dans [Duran 2019, p. 15-16], étant entendu que certaines références spécifiques seront citées plus loin.

4. Les *Mathematische Annalen* ont déjà été identifiées comme lieu privilégié du développement de la géométrie algébrique à la fin du XIX^e siècle. Voir [Rowe et Tobies 1990, p. 37-46], où sont par ailleurs listés et brièvement décrits les auteurs les plus prolifiques de la revue ainsi que les principaux sujets qui y sont abordés. La géométrie algébrique est l'un d'eux, mais ses différents sous-domaines n'y sont pas distingués.

5. Exclure le sujet des surfaces quadriques, parfois vu comme plus élémentaire, permet d'éluder un problème d'interférences entre géométries algébrique et différentielle : le sous-titre du paragraphe qui leur est dévolu indique qu'y sont classifiés les articles abordant ces surfaces tant par des considérations algébriques que différentielles ; réciproquement, la section de géométrie différentielle renvoie à ce paragraphe pour le sujet du traitement différentiel des quadriques.

plus, afin de me restreindre à une approche textométrique monolingue, seuls les textes écrits en allemand ont finalement été retenus.

Ce corpus, si spécifique qu'il soit, présente l'avantage d'être constitué d'un nombre de mots suffisamment important pour donner sens aux différents calculs statistiques auxquels nous aurons affaire tout en restant raisonnable quant au nombre de textes qui le composent, facilitant ainsi le contrôle des résultats de cette première étude textométrique. La restriction à la langue allemande vise quant à elle à éliminer la question de la mise en correspondance de langues différentes, et donc à limiter pour le moment la complexité de la tâche.

Par la manière dont il a été constitué, le corpus permettra d'apporter un éclairage nouveau sur l'histoire traditionnelle du sujet des surfaces algébriques dans la deuxième moitié du XIXe siècle [6]. En effet, mis à part le thème des modèles de surfaces et celui (non indépendant) des travaux des années 1870 et 1880 de Felix Klein et ses collaborateurs sur la surface de Kummer et les complexes de droites, abordés dans des recherches historiques plutôt récentes [7], peu de contributions allemandes de l'époque sont présentes dans cette histoire, apparemment victimes du point de fuite que constitue la géométrie italienne de la fin du siècle : le récit usuel commence en général par mentionner les recherches sur les surfaces cubiques lancées en 1849 par Arthur Cayley et George Salmon, et suivies par celles de divers géomètres comme Jacob Steiner, Ludwig Schläfli, James Joseph Sylvester ou Luigi Cremona. Sont ensuite mis au centre de la scène des travaux de Max Noether de 1870 et 1875 (précédés de quelques contributions de Cayley et d'Alfred Clebsch, auxquels est parfois ajouté Hieronymus Zeuthen) sur la théorie générale des surfaces et la considération d'intégrales doubles sur celles-ci, travaux repris et étendus une dizaine d'années plus tard par Émile Picard dans sa « théorie transcendante » des surfaces. La main semble ensuite passer entièrement du côté de l'« école de géométrie algébrique italienne », autour de 1890 [8].

L'étude entreprise ici n'a cependant pas pour seul but de (commencer à) compléter ce récit, en mettant en évidence des éléments encore mal

6. Voir par exemple [Dieudonné 1974 ; Gray 1989 ; Kolmogorov et Yushkevich 1981/1996 ; Houzel 2002].

7. Parmi les dernières références en la matière, voir [Rowe 2017 ; Volkert 2019 ; Barrow-Green 2020] ainsi que les différentes contributions d'un numéro spécial de *Mathematical Intelligencer* de 2017, reprenant certaines des interventions de [Gray et al. 2015].

8. Ces différents épisodes ont eux-mêmes été récemment documentés au regard de problématiques historiques propres. Outre les références données *supra*, voir par exemple [Brigaglia, Ciliberto et Pedrini 2004 ; Lê 2015 ; Schappacher 2015 ; Brigaglia 2016 ; Casnati et al. 2016].

connus historiquement, comme des thèmes de recherche (la dite « nouvelle géométrie synthétique [9] » et les représentations de surfaces spéciales) ou des textes clés (dont un article de Clebsch [1869] sur ces représentations). Au delà de tels éléments, dont la mise au jour dépend en fait surtout de la systématisation du processus de formation du corpus et de sa lecture, seront en effet présentés plusieurs types de résultats et de pistes de recherche en histoire des mathématiques liés à l'utilisation des outils textométriques et que je crois être originaux : richesse et dynamisme lexicaux, colorations sémantiques de termes choisis, études comparatives du lexique et des parties du discours de groupes de textes, division d'un corpus par classification lexicale, discutée notamment vis-à-vis du point de vue des rubriques disciplinaires du *Jahrbuch* et de celui des réseaux de citations.

Plus précisément, nous verrons par exemple que deux mots comme « géométrique » et « algébrique » ne sont pas associés aux mêmes réseaux sémantiques au sein du corpus, de même que le mot « équation », lorsqu'on l'évalue comparativement à un autre corpus, celui de la théorie des invariants. La confrontation avec ce corpus algébrique fera aussi voir des différences notables sur la nature des verbes utilisés de part et autre, ainsi qu'une répartition inéquitable des types de parties du discours, dévoilant ainsi des modes d'écriture propres à chacun des corpus. Enfin, le regroupement en classes lexicales des textes relatifs aux surfaces algébriques, allié avec les possibilités de mise en évidence des termes caractéristiques de ces classes, donnera un aperçu sur la structuration du sujet en différentes sous-thématiques, se superposant en grande partie avec celle obtenue par le point de vue des réseaux de citations.

Mais décrivons d'abord brièvement quelques généralités relatives au corpus construit. Choisir l'allemand comme langue de référence a pour effet d'écarter 16 textes sur les surfaces algébriques écrits dans d'autres langues et laisse ainsi 75 articles, dont 51 relèvent des surfaces spéciales et 24 des surfaces générales [10]. On compte en tout 26 auteurs, parmi lesquels 14 publient uniquement en surfaces spéciales, 4 en surfaces générales et 8 dans les deux. Ces auteurs contribuent à différents degrés au corpus : neuf d'entre eux y signent une seule publication, tandis que les deux plus prolifiques, Rudolf Sturm et Alfred Clebsch, y apportent respectivement

9. Cette expression est le titre d'une des rubriques disciplines du *Jahrbuch*. Les textes du notre corpus relevant de cette thématique seront discutés plus loin.

10. Le découpage entre surfaces générales et surfaces particulières n'a pas été retenu comme critère d'étude *a priori*, et tous les calculs statistiques seront donc aveugles en la matière. Nous verrons toutefois comment ce découpage se retrouve (très partiellement) dans l'interprétation de certains résultats donnés par l'analyse textométrique.

7 et 6 articles [11].

Parmi les 16 textes écartés, 10 sont écrits en français, 3 en italien et 3 en anglais. Ces derniers sont tous signés de la main d'Arthur Cayley et ont été publiés en 1871. Les trois articles italiens sont de Federigo Enriques et de Guido Castelnuovo, et datent d'entre 1894 et 1897. Ces deux auteurs publient aussi ensemble un article en français en 1897. Outre une contribution de Corrado Segre de 1884, une autre de George Humbert de 1894 et une dernière d'Ernest de Jonquières de 1869, toutes les autres publications en français ont été écrites par Hieronymus Zeuthen, et sont parues entre 1871 et 1876.

Les dates de la toute fin du siècle associées aux publications de Enriques, Castelnuovo et Humbert sont significatives, et, en négatif, mettent en perspective la répartition chronologique du corpus allemand. Celle-ci peut en effet se décrire en deux vagues principales : la première s'étend de 1869 à 1875 environ et regroupe 35 articles (avec un pic de 17 articles publiés entre 1871 et 1872), tandis que la seconde, plus longue et moins haute, couvre les années 1878 à 1890 avec 39 articles — il ne reste qu'un seul article en allemand publié après cette deuxième vague, en 1897. Les travaux allemands sur le sujet des surfaces algébriques s'effacent ainsi des *Mathematische Annalen* lors de la dernière décennie du XIXe siècle, au profit de quelques contributions italiennes et françaises [12].

Ces premières informations au sujet du corpus de référence étant présentées, passons maintenant à l'approche textométrique. L'outil informatique retenu est TXM, un logiciel libre développé par Serge Heiden, [Heiden, Magué et Pincemin 2010], dont l'utilisation nécessite une préparation technique des textes visant à être analysés. Cette préparation est l'objet d'une première section, où sont notamment explicités tous les choix qui ont été faits à cette étape et qui ont une incidence sur la manière d'interpréter les résultats [13]. Ceux-ci sont présentés dans

11. Ces nombres renvoient à deux cas de figure distincts : alors que Sturm (1841-1919) a pu publier dans les *Mathematische Annalen* tout au long de la période qui nous intéresse ici, Clebsch (1833-1872) meurt brutalement au tout début de celle-ci.

12. Cette constatation conforte *a priori* la partie de l'histoire usuelle selon laquelle la géométrie algébrique italienne (accompagnée de quelques contributions françaises) prend le relais à partir du début des années 1890. Soulignons toutefois que l'image dégagée ici reste très partielle, puisque construite uniquement à partir des surfaces algébriques dans les *Mathematische Annalen*, au sens défini plus haut. Par ailleurs, la grosse moitié de notre corpus publiée entre 1878 et 1890 (dont on verra qu'elle contient bien d'autres articles que ceux consacrés à la question des modèles de surfaces) contrebalance de fait le relatif vide historiographique pour toute la période qui suit les contributions de Noether.

13. De telles précautions et questionnements relatifs à une approche numérique d'un corpus en histoire des mathématiques sont aussi explicités, dans d'autres cas,

les trois sections suivantes, dévolues respectivement à trois manières de mettre en lumière les données textométriques. Ainsi, la deuxième section étudie à grands traits quelques aspects du vocabulaire du corpus sans le confronter sérieusement à un autre, ce que fait la troisième section en introduisant comme corpus de comparaison celui de la théorie des invariants. La quatrième section, quant à elle, revient au corpus des surfaces algébriques seul et en propose une analyse dont le point de départ est une partition des textes du corpus en six classes lexicales.

1. Prolégomènes techniques

La fonctionnalité la plus élémentaire du logiciel TXM est de lister tous les mots d'un corpus avec leurs fréquences d'apparition [14]. À chacun de ces mots est attachée une étiquette indiquant sa nature grammaticale et son lemme, c'est-à-dire, essentiellement, l'entrée qui correspondrait au mot en question dans un dictionnaire. C'est sur la base de ces éléments que s'appuient ensuite tous les autres calculs statistiques proposés par TXM, et qui permettent l'analyse textométique à proprement parler.

Or, cette liste dépend directement de la manière dont les constituants du corpus ont été traités avant d'être livrés au logiciel. En effet, chacun d'eux doit donner lieu à un fichier informatique contenant le texte qu'il représente dans un format adéquat. Celui qui a été utilisé ici est le simple format `txt` ; chaque fichier a été construit sur la base des données de texte mises à disposition par différents sites d'archives ouvertes et générées par un procédé de reconnaissance optique de caractères. Afin de garantir un décompte correct des mots par TXM ainsi que leurs bonnes lemmatisation et reconnaissance grammaticale, une étape de vérification de ces données a ainsi été nécessaire pour éliminer toute coquille [15]. Mais il est également nécessaire d'intervenir pour standardiser les fichiers `txt`, en raison cette fois de la nature même des textes étudiés : des textes mathématiques techniques, écrits dans la langue allemande de la fin du XIX^e siècle.

On y trouve en effet quelques rares mots dont l'orthographe n'est pas encore tout à fait stabilisée, parfois même entre plusieurs textes d'un même auteur : c'est le cas de *Mannigfaltigkeit* et de *Mannichfaltigkeit*,

dans [Guilbaud et al. 2014 ; Goldstein 2018].

14. J'adopte dans cet article la terminologie des textomètres : la « fréquence » d'un mot désigne le nombre absolu de ses occurrences dans un corpus donné.

15. La qualité des données de texte ainsi récupérées est très fluctuante, notamment (semble-t-il) à cause de la présence des nombreux symboles mathématiques qui gênent la bonne reconnaissance optique des caractères des textes originaux. La vérification manuelle des données a été faite sur l'intégralité des textes du corpus.

qui désignent pourtant le même objet mathématique. Dans la perspective d'étudier la distribution lexicale au sein du corpus, j'ai choisi de remplacer l'une des deux variantes par l'autre, permettant ainsi à TXM de les compter comme deux occurrences identiques [16]. Un autre aspect relatif à la langue allemande de l'époque est que de nombreuses conventions orthographiques, adoptées cette fois uniformément par tous les auteurs étudiés, ne sont pas les mêmes que celles d'aujourd'hui : par exemple, « Curve », « nöthig » et « existiren » s'écrivent maintenant « Kurve », « nötig » et « existieren ». La reconnaissance des lemmes et l'analyse grammaticale de TXM se basant sur un dictionnaire actuel, tous les textes du corpus ont été convertis aux normes orthographiques actuelles [17].

Une autre difficulté dans la préparation des fichiers informatiques a été de gérer les formules mathématiques dont regorgent les textes originaux. Outre le problème de transcrire de manière standardisée ces formules au format `txt`, celui de les interpréter en termes de mots afin de les séparer (ou de les rassembler) et de les compter s'est révélé particulièrement épineux. Par exemple, une formule apparemment simple comme « $f_0(x) = 2x + 1$ » montre déjà que la question de déterminer de combien de mots (mathématiques) elle se compose est loin d'être évidente. Les listes de symboles, comme « P, Q », sont elles aussi délicates de ce point de vue : on trouve ainsi, parfois au sein d'un même paragraphe, des occurrences comme « les points P, Q » et « le couple P, Q » qui pourraient appeler à deux types de comptages de mots ; mais la situation se complique encore quand le ou les objets ainsi désignés ne sont plus explicités, comme dans l'expression « à P, Q est associée une droite ».

Devant la multiplicité de telles situations (souvent bien plus complexes) et les difficultés techniques que présenterait la prise en compte des formules [18], j'ai opté pour une solution radicale consistant à effacer tout le contenu mathématique de ces formules tout en gardant une trace de leur existence. Plus précisément, chaque formule mathématique a été

16. On pourrait bien entendu vouloir au contraire conserver ces variantes orthographiques et voir si elles sont susceptibles d'apporter des informations pertinentes. J'ai décidé ici de ne pas suivre cette voie, qui ne me paraissait pas prometteuse dans les cas observés.

17. Il s'agit bien d'une sorte de substitution systématique de certaines lettres ou groupes de lettres par d'autres, et pas, par exemple, d'une transformation plus profonde des racines des mots. Effectuer cette substitution a été grandement facilité par le script développé et mis librement à disposition par Bryan Jurish, `http://www.deutschestextarchiv.de/demo/cab/file`. Voir [Jurish 2011] pour plus d'informations techniques à ce sujet.

18. Il est possible de soumettre au logiciel TXM des textes en format enrichi XML, ce qui pourrait permettre d'intégrer les formules mathématiques. Je n'ai cependant pas cherché à creuser cette piste à ce stade de mes recherches.

remplacée par un unique symbole $*$ lorsqu'elle se trouve en plein texte, et par un symbole # lorsqu'elle est centrée hors paragraphe ; cela vaut aussi pour les énumérations, mais j'ai tout de même continué à distinguer deux symboles mathématiques séparés par une virgule lorsque celle-ci est nécessaire pour respecter la syntaxe d'une phrase. Prenons un exemple fictif de phrase originale [19] :

> Jede Gruppe von Curven mit der Degeneration u, v besitzt $2u + 1$ Punkte l, $v^2(2v^2 - 3v + 2)$ Seiten l', und
>
> $$v(2u + 1)(2v^2 - 3v + 2)$$
>
> Ebenen E.

Le traitement de cette phrase la transforme dans le fichier `txt` en la suivante :

```
Jede Gruppe von Kurven mit der Degeneration * besitzt *
Punkte *, * Seiten *, und # Ebenen *.
```

On remarquera entre autres que le groupe de symboles l, $v^2(2v^2 - 3v + 2)$ n'a pas été fondu en une seule $*$ car la première lettre, l, est syntaxiquement attachée à « Punkte », tandis que l'expression avec les v dénombre les « Seiten » qui suivent.

Il est clair que ce choix ampute les textes du corpus d'une grande partie de leur contenu. Néanmoins, la focalisation ainsi forcée sur l'information mathématique relayée par les mots du langage naturel, ainsi que la considération du nombre même de formules incarnées par des symboles $*$ et # donnent déjà des résultats intéressants qui seront exposés dans la suite.

Du reste, le passage au format `txt` gomme d'autres types de données textuelles qui ne relèvent pas directement du lexique et qu'il faut donc accepter de perdre dans le traitement statistique, vu les conventions adoptées. C'est le cas des dispositifs typographiques particuliers (indentations exceptionnelles, utilisation d'italique ou modification de l'interlettrage), de la position exacte des notes de bas de page, des dispositions tabulaires ou encore des figures [20].

19. Cet exemple est inspiré de [AFFOLTER 1887, p. 17], des modifications ayant été apportées pour montrer d'un coup plusieurs types de problèmes. La traduction est la suivante : « Chaque groupe de courbes de dégénération u, v possède $2u + 1$ points l, $v^2(2v^2 - 3v + 2)$ côtés l' et $v(2u + 1)(2v^2 - 3v + 2)$ plans E. »

20. Le contenu des notes de bas de page a bien sûr été conservé, mais a systématiquement été placé dans le corps du texte, juste après la fin des phrases dans lesquelles elles sont appelées. Un interlettrage large est employé systématiquement dans les *Mathematische Annalen* pour les noms de personnes : « Clebsch » s'y trouve ainsi écrit « C l e b s c h ». C'est aussi le cas pour les énoncés de certains théorèmes, mais seulement dans les deux premiers volumes, l'italique devenant ensuite la norme.

Une fois ce formatage effectué, les fichiers `txt` sont livrés à TXM. Comme indiqué précédemment, deux opérations essentielles sont alors réalisées. D'une part, la nature grammaticale de chaque mot est identifiée et distribuée selon certaines sous-catégories : sont par exemple distingués les noms communs, les noms propres, les articles, les adjectifs attributifs, les adjectifs prédicatifs et adverbiaux, les verbes pleins, modaux et auxiliaires sous forme finie, infinitive ou participiale, les différents types de pronoms, particules et autres conjonctions [21]. D'autre part, les mots du corpus sont tous associés à leur lemme de référence : forme infinitive pour les verbes, forme singulière nominative pour les noms communs, forme positive prédicative pour les adjectifs, etc. [22].

Le traitement initial du corpus par TXM étant effectué, tous les mots, lemmes et parties du discours sont listés et énumérés. Il est alors possible de parcourir le lexique du corpus, de procéder à diverses recherches, mais aussi de consulter les textes eux-mêmes *via* une édition construite par le logiciel, offrant une navigation efficace entre ceux-ci et les listes de résultats.

2. Premier aperçu textométrique

Le corpus des surfaces algébriques compte 632 926 mots pour 11 849 formes distinctes. Comme le montre la table 1, le mot le plus fréquent est le substitutif $*$ des formules mathématiques non centrées, employé 40 093 fois. Abondent également les signes de ponctuation, les formes *die*, *der* et *den* (qui peuvent être des articles définis ou des pronoms relatifs), le symbole # des formules centrées, la forme conjuguée *ist* de l'auxiliaire *sein* (« est », « être [23] »), ainsi que plusieurs conjonctions, prépositions et adverbes. Les premiers mots pleins sont des substantifs qui, sans

21. Toutes ces catégories grammaticales sont celles de TreeTagger, le logiciel d'étiquetage morpho-syntaxique sous-traité par TXM. La liste des 54 catégories de la langue allemande utilisées est décrite sur la page https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/. Les verbes reconnus comme auxiliaires sont *sein*, *haben* et *werden* ; les modaux sont *wollen*, *sollen*, *mögen*, *können*, *dürfen*, *müssen* ; les verbes pleins sont tous les autres verbes.

22. Le traitement automatique des lemmes et catégories grammaticales n'est pas exempt de failles, en particulier car la présence des symboles $*$ et # a tendance à entraver l'analyse morpho-syntaxique des phrases. Là encore, un travail de vérification manuel a été conduit sur tout le corpus pour contrôler, voire corriger, l'étiquetage automatique.

23. Dans tout ce qui suit, les traductions ne seront écrites qu'une fois, lors de la première occurrence du mot allemand correspondant, sauf pour les mots transparents qui ne seront pas traduits. En outre, tous ces mots seront donnés avec leur orthographe actuelle.

grande surprise, reflètent la thématique générale du corpus en renvoyant aux notions de base de la géométrie dans l'espace : *Fläche(n)*, *Ordnung*, *Punkte*, *Kurve*, *Geraden*, *Ebene* (« surface(s) », « ordre », « points », « courbe », « droites », « plan »). Les hapax, c'est-à-dire les mots qui surviennent une et une seule fois à travers tout le corpus, sont quant à eux au nombre de 4 295.

Mots	Fréquence	Mots	Fréquence	Mots	Fréquence
*	40 093	ist	5 733	;	3 652
,	38 649	mit	5 167	sind	3 431
.	30 132	auf	5 121	zu	3 379
die	21 684	Ordnung	4 664	(	3 378
der	20 324	so	4 417	dass	3 343
und	11 132	)	4 403	des	3 125
in	10 238	sich	4 318	Kurve	2 962
von	8 534	welche	4 257	Geraden	2 941
den	6 328	:	4 039	zwei	2 740
eine	6 205	Punkte	4 001	aus	2 337
durch	6 151	Die	3 873	Ebene	2 259
#	5 904	einer	3 745	Flächen	2 245
Fläche	5 790	man	3 732	ein	2 240

TABLE 1 – Les premiers mots du lexique, classés par ordre décroissant de fréquence d'apparition.

Pour mettre ces premiers nombres en perspective, notons qu'un texte comme *Also sprach Zarathustra* (1883-1885) de Friedrich Nietzsche est formé d'environ 109 777 mots pour 11 809 formes distinctes, dont 6 668 hapax, ou que le *Kapital* (1867) de Karl Marx compte 947 372 mots répartis en 52 504 formes distinctes, dont 27 629 hapax (voir la table 2) [24]. Ces chiffres montrent que le lexique de notre corpus mathématique, comportant peu de formes distinctes et peu d'hapax par rapport au nombre total de mots, est peu varié en comparaison avec *Zarathustra* et le *Kapital*, et ce, malgré la multiplicité des auteurs participant au corpus. Cette pauvreté lexicale, qui reflète sans doute la nature technique des textes considérés ainsi qu'une certaine standardisation de ceux-ci (les mots connaissant un fort taux de réemploi), n'est d'ailleurs pas le

24. Les données pour ces deux textes ne sont qu'approximatives car, si les fichiers sources `txt` de ces derniers (qui proviennent du site du projet *Deutsches Textarchiv*, http://www.deutschestextarchiv.de) semblent tout à fait corrects du point de vue orthographique, ils comportent tout de même quelques mots parasites, comme les numéros de page ou le paratexte éditorial, que je n'ai pas cherché à éliminer.

seul fait de la prolifération des symboles $*$ et # : même en les ôtant du décompte, les rapports entre nombre de mots, formes distinctes et hapax restent du même ordre de grandeur [25].

	Mots	Formes distinctes	Hapax
Surfaces algébriques	632 926	11 849 (1,9 %)	4 295 (0,7 %)
Zarathustra	109 777	11 809 (10,8 %)	6 668 (6,1 %)
Das Kapital	947 372	52 504 (5,5 %)	27 629 (2,9 %)

TABLE 2 – Comparaison des nombres de mots, de formes distinctes et d'hapax entre le corpus des surfaces algébriques, *Also sprach Zarathustra* et *Das Kapital*. Les pourcentages indiqués indiquent les ratios relatifs aux nombres de mots totaux.

Le vocabulaire du corpus des surfaces algébriques n'est cependant ni figé dans le temps, ni réduit à celui d'un petit nombre de textes qui suffiraient à le décrire intégralement. S'il est en effet clair que certains termes sont universellement employés (comme les mots les plus fréquents relevés précédemment), le graphe présenté en figure 1 montre que le nombre cumulé de formes distinctes croît tout au long de la période considérée, moins intensément mais tout aussi régulièrement que le nombre total de mots. Certaines fonctionnalités de TXM permettent en outre de repérer des mots qui apparaissent (ou disparaissent) à tout moment, sans qu'ils soient pour autant des hapax. Les exemples (de faible fréquence) présentés en figure 2 montrent d'ailleurs que les termes qui apparaissent ne sont pas nécessairement des néologismes créés à l'occasion ; il peut aussi s'agir de mots non originaux dont l'utilisation reflète par exemple l'activation d'une certaine thématique de recherche ou l'adoption d'une méthode particulière. Bien que relativement pauvre au sens expliqué ci-dessus, le lexique du corpus n'en est donc pas moins tout à fait dynamique.

Il est bien sûr possible d'interroger le corpus plus finement pour dépasser ces chiffres globaux, en s'intéressant à des mots ou des groupes de mots donnés. L'étiquetage grammatical opéré par TXM donne par exemple immédiatement accès à la liste des 206 différents noms propres, apparaissant en tout à travers 1208 occurrences. Sans entrer dans les détails [26], signalons juste que les noms les plus fréquents sont ceux

25. Bien que cela dépasse largement le cadre de cet article, il serait intéressant de définir et d'évaluer une notion de richesse lexicale tenant compte du contenu des formules mathématiques.

26. Une étude des noms propres dans le corpus formé des *Œuvres complètes* de Charles Hermite est faite dans [GOLDSTEIN 2012].

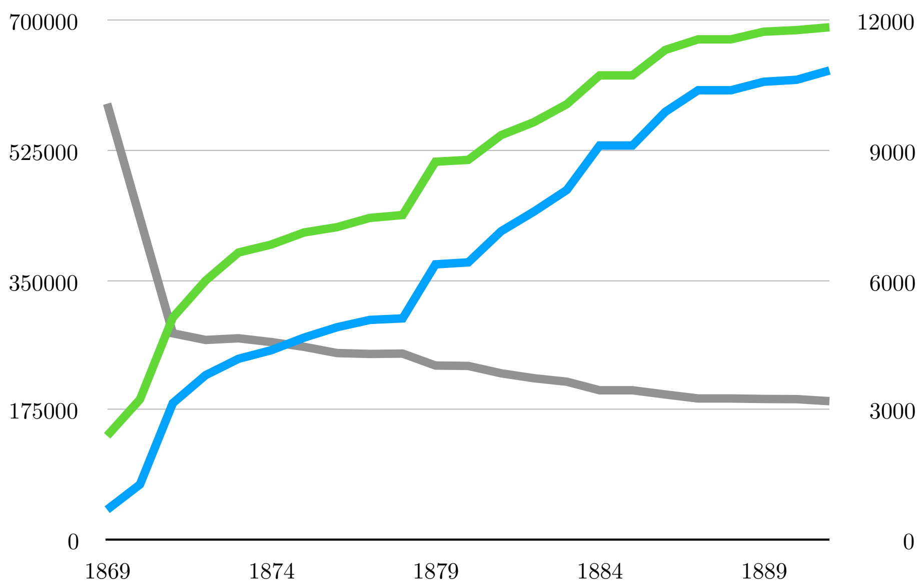

FIGURE 1 – Nombres cumulés de mots (en bleu, axe de gauche) et de formes distinctes (en vert, axe de droite), et rapport de ces derniers sur les premiers (en gris). Les années 1891 à 1896, sans publication, n'ont pas été représentées.

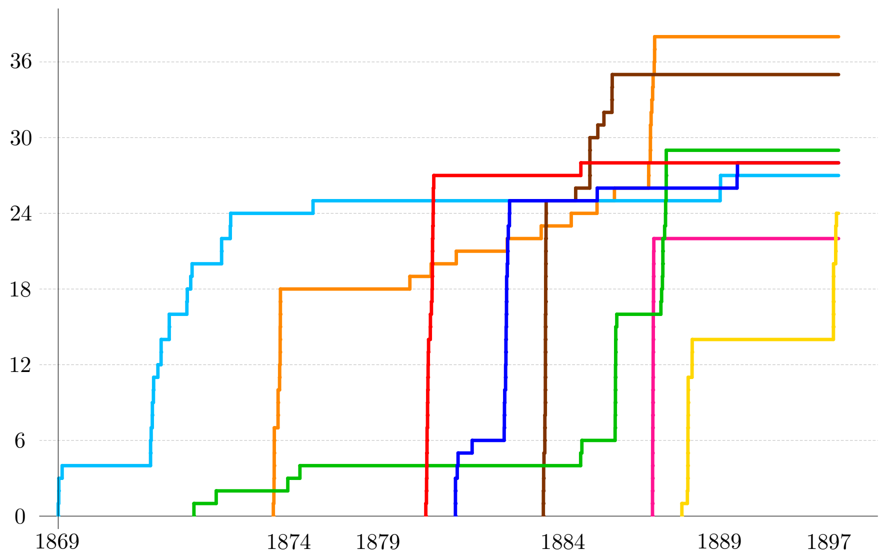

FIGURE 2 – Nombres cumulés de certains lemmes au fil des textes, eux-même ordonnés chronologiquement. ■ : *abbildbar*. ■ : *Degeneration*. ■ : *Modell*. ■ : *Thetafunktion*. ■ : *zyklisch*. ■ : *Potenzreihe*. ■ : *ideell*. ■ : *Hauptpunkt*.

de Clebsch (114 occurrences), Cayley (66 occurrences), Cremona (60 occurrences), Zeuthen (45 occurrences) puis de Schläfli et de Sturm (41 occurrences chacun). On remarquera au passage que le trio de tête est celui que Wilhelm Fiedler décrivait en 1867 comme constituant « le grand C qui, surtout en termes de productivité, marche actuellement au sommet de l'Europe mathématique dans le domaine de la géométrie analytique ou de l'analyse géométrique (au sens le plus large) [27] ».

Pour prendre un autre exemple, plus poussé, regardons comment les mots « géométrique(s) », « géométriquement », « algébrique(s) », « algébriquement » interviennent dans le corpus. En allemand, il s'agit de s'intéresser aux mots dont le lemme est *geometrisch* ou *algebraisch* : on en compte 158 pour le premier (dont 51 *geometrische*, 44 *geometrischen* et 35 *geometrisch*) et 416 pour le deuxième (dont 177 *algebraische*, 149 *algebraischen*, 52 *algebraisch* et 29 *algebraischer*).

Afin d'identifier des thématiques associées à ces mots, une fonction de TXM permet d'en lister les cooccurrents spécifiques, c'est-à-dire les termes qui se démarquent par leur présence singulièrement élevée dans les voisinages des mots donnés [28]. Ces cooccurrents (voir la table 3) font deviner des associations sémantiques intéressantes, qu'il convient de confirmer et d'affiner en examinant de manière précise comment elles s'ancrent effectivement dans les textes. Une manière efficace de procéder consiste à utiliser les concordanciers produits par TXM, qui sont les listes de toutes les occurrences d'un mot (ou d'un lemme) replacées dans leur contexte immédiat, assorties de fonctions de tris divers et de retour plus complet aux textes d'origine.

Cette étude montre que la plupart des cooccurrents de *geometrisch* d'une part, et de *algebraisch* d'autre part, sont en fait aussi des cooccurrents les uns des autres. Ainsi, quand il apparaît aux côtés de *geometrisch*, l'adjectif *rein* est systématiquement employé de manière adverbiale (« purement ») et en précise le sens ; il est alors principalement associé à d'autres cooccurrents comme *Weg* et *Hilfsmittel*, et qualifie

27. Extrait d'une lettre du 13 mars 1867 de Wilhelm Fiedler à Luigi Cremona : « Mr. Cayley ist mit Ihnen, werthester Freund, u. Clebsch der dritte in dem grossen C, welches jetzt vorzugsweise productiv an der Spitze des mathematischen Europa in dem Gebiete der analytischen Geom. oder geometrischen Analysis (im weitesten Umfange gedacht) marschirt. » [Israel 2017, p. 653].

28. Il s'agit de distinguer, parmi tous les mots qui composent tous les voisinages (de taille arbitrairement choisie) des mots donnés, ceux qui surviennent remarquablement plus que les autres, et de croiser cela avec leur fréquence d'apparition propre. La mesure précise de ce phénomène est basée sur des calculs d'*indices de spécificité*. Des explications plus détaillées sur la notion de spécificité seront données dans la prochaine section, autour d'un exemple un peu plus parlant.

geometrisch	*algebraisch*
rein	algebraisch
Ort	Minimalfläche
Interpretation	Evolute
Weg	Satz
Bedeutung	Developpable
Lösung	einschreiben
Problem	Funktion
analytisch	vorlegen
Hilfsmittel	ich
leicht	Minimalkurve
Abbildung	Theorie

TABLE 3 – Les premiers lemmes cooccurrents des lemmes *geometrisch* et *algebraisch*, classés par indice de spécificité décroissant.

donc de « purement géométriques » des « cheminements », des « outils » proposés ou cités par les auteurs, et par ailleurs souvent aussi décrits comme « simples » (*leicht*). L'adjectif *analytisch* n'a pas tout à fait la même fonction : au contraire de *rein*, il ne précise pas *geometrisch* mais s'y oppose, dans des expressions comme « des considérations analytiques et géométriques » où les deux adjectifs font référence à deux types de démarches : « géométrique » et « analytique » renvoient alors implicitement à la géométrie pure et à la géométrie analytique. Outre l'association avec des noms d'objets mathématiques comme *Ort* et *Abbildung* (« lieu », « représentation »), on remarquera aussi que *geometrisch* est lié à plusieurs substantifs qui pointent vers le champ lexical de l'interprétation et de la signification (*Interpretation*, *Bedeutung*, etc.). Dans ces cas-là, *geometrisch* change sensiblement de coloration, puisqu'il s'agit la plupart du temps d'interprétations géométriques de calculs ou d'équations : il ne s'agit ainsi plus d'opposer géométrie pure et géométrie analytique, mais d'affecter un sens géométrique à des formules mathématiques.

La situation est toute différente pour le lemme *algebraisch*. Les concordanciers montrent déjà de manière évidente que l'immense majorité des occurrences du lemme proviennent de deux vastes textes (qui sont en fait deux parties d'un même mémoire) de Sophus Lie sur les surfaces minimales algébriques [LIE 1879a,b], alors que *geometrisch* n'était pas employé par un auteur en particulier. L'adjectif *algebraisch* sert alors à qualifier des objets (*Minimalfläche*, *Evolute*, *Developpable* : « surface minimale », « développée », « développable ») ainsi que des théorèmes qui les impliquent, et dont les versions générales (c'est-à-dire non nécessairement algébriques) existent et relèvent de la géométrie différentielle [29]. En outre, l'emploi intense et inlassable de *algebraisch* par Lie dans des syntagmes qui impliquent plusieurs des termes évoqués à l'instant explique que ce lemme soit un de ses propres cooccurrents. Notons aussi l'exception notable de la liste des cooccurrents qu'est le substantif *Funktion*, celui-ci étant souvent qualifié d'algébrique dans des travaux de Noether. En revanche, la liste complète des cooccurrents de *algebraisch* ne contient aucun mot qui renverrait au réseau sémantique de l'interprétation, offrant encore une fois un net contraste avec le cas de *geometrisch*.

De telles colorations sémantiques de *algebraisch* et *geometrisch* laissent deviner en creux des différences de représentation et de mobilisation des domaines disciplinaires sous-jacents qui mériteraient d'être analysées plus en profondeur ou confrontées à d'autres situations [30]. Je me contenterai ici simplement de relever le potentiel heuristique et exploratoire qu'offrent les outils textométriques présentés dans cette section. Ces outils seront à nouveau utilisés dans les deux sections suivantes, consacrées à des études contrastives du corpus des surfaces algébriques.

3. Une analyse contrastive exogène : surfaces vs. invariants

Cherchant un corpus auquel comparer celui des surfaces algébriques, et donc de nature et de taille analogues à celles de ce dernier, j'ai opté pour celui formé de tous les textes relevant de la théorie des invariants, repérés comme dans le cas des surfaces à l'aide du *Generalregister* des *Mathematischen Annalen* pour les années 1869 à 1898. Le corpus ainsi obtenu est composé de 105 articles classés en « Théorie des invariants. Généralités » pour 63 d'entre eux et en « Théorie des invariants de formes particulières » pour les 42 restants, deux paragraphes de la section d'algèbre [31], elle-même contenue dans le chapitre « Arithmétique et algèbre » du registre. Ces articles sont écrits par 39 auteurs, dont 8 sont aussi des contributeurs à la théorie des surfaces : Alexander Brill, Alfred Clebsch,

29. Nous verrons en section 4 que ces textes de Lie sont bel et bien des cas tout à fait isolés au sein du corpus des surfaces algébriques, à la fois en termes de réseaux de citations, de classes lexicales et de classification disciplinaire.

30. Par exemple, le fait que *geometrisch* active le réseau sémantique de l'interprétation pourrait être lié au choix de s'être restreint aux *Mathematische Annelen*, revue diffusant notamment les articles de mathématiciens proches de Clebsch et de Klein.

31. Tout comme pour les surfaces, le *Generalregister* distingue donc le général et le particulier au sein du sujet des invariants. Comme précédemment, l'ensemble des publications relevant de l'une ou l'autre de ces thématiques sera considérée comme un unique corpus.

Paul Gordan, Max Noether, Axel Harnack, Felix Klein, Hermann Thieme et Aurel Voss [32].

Reprenons déjà les données générales que nous avons présentées précédemment pour les surfaces : 632 926 mots, 11 849 formes distinctes, 4 295 hapax. Les données correspondantes pour le corpus des invariants — formaté et corrigé de la même manière que celui des surfaces — sont respectivement égales à 460 327, 11 145 et 4 249. Les différents rapports entre ces nombres indiquent déjà que les deux corpus ont un profil analogue en comparaison avec les textes non mathématiques que sont *Zarathustra* et le *Kapital*. Il est de plus à fois curieux et remarquable que les nombres de formes distinctes et d'hapax soient aussi proches malgré la différence que l'on peut voir au niveau des nombres de mots ; sans chercher à en extrapoler une quelconque loi, on notera simplement que les deux corpus disposent donc de viviers lexicaux de tailles identiques mais que celui des surfaces connaît un taux de réemploi plus élevé. Deux autres points de ressemblance entre les corpus, enfin, sont la croissance ininterrompue de leur lexique entre 1869 et 1898 (que ce soit du point de vue du nombre de mots ou de formes distinctes), ainsi que leur dynamisme, tel que mesuré par l'apparition et la disparition de termes au cours de cette période — je ne redonne pas ici les figures et détails analogues à ceux présentés dans le cas des surfaces.

Les formules mathématiques centrées, en revanche, marquent une forte différence entre surfaces et invariants : si les symboles $*$, rapportés aux nombres de mots, sont à peu près équivalents pour les deux corpus (40 093 pour les surfaces contre 30 999 pour les invariants), on observe un déséquilibre très net sur les symboles #, puisque les invariants en comptent 10 081 alors que les surfaces n'en totalisent que 5 904. Cette constatation sur les formules centrées peut aussi être vue en mettant en rapport les nombres totaux de pages (d'origine) des corpus et les nombres de mots relevés par TXM : si le corpus des surfaces compte en moyenne 340 mots par page, celui des invariants en affiche seulement 230. Cette différence s'explique par la mise en forme préliminaire des fichiers `txt`, toutes les formules centrées, parfois extrêmement volumineuses, ayant été remplacées par les seuls symboles #, faisant ainsi fondre la quantité de mots par page pour les textes sur les invariants, qui sont ceux qui en comptent le plus.

Un tel résultat peut paraître banal, voire évident (quoi qu'il ait été

32. À l'exception peut-être de Thieme, tous ces doubles contributeurs sont liés de près aux cercles de Clebsch et de Klein. Les liens sociaux de Klein (et, dans une moindre mesure, de Clebsch) sont exposés dans [Tobies 2019]. Pour quelques informations au sujet de Thieme, voir [Hilbert 1899/2015, p. 216-217].

ici proprement quantifié), à quiconque ayant déjà lu un certain nombre de textes de la fin du XIX^e siècle relevant de la théorie des invariants et de la théorie des surfaces. Dans la suite de cette section, il sera toutefois complété en suivant deux pistes offertes par les outils textométriques. La première est l'étude comparative des cooccurrents du même lemme *Gleichung* (« équation ») dans chacun des corpus, dont les premiers sont donnés dans la table 4. Quant à la seconde, elle vise à mettre au jour quelques caractéristiques d'écriture dans chaque corpus, en se basant sur la distribution des parties du discours qui y sont employées.

Surfaces		Invariants	
Cooccurrent	Spécificité	Cooccurrent	Spécificité
Gleichung	74	Wurzel	46
Form	53	determinierend	36
Wurzel	28	genügen	34
setzen	27	Lösung	24
stellen	24	befriedigen	21
Elimination	23	Auflösung	15
eliminieren	20	Seite	13
Faktor	19	Elimination	11
befriedigen	18	ergeben	9
erhalten	17	links	9
Koordinate	17	bestehen	8
homogen	15	fünft	8
genügen	15	rechts	8

Table 4 – Les lemmes (correspondant à des mots pleins) cooccurrents au lemme *Gleichung* les plus spécifiques.

Commençons avec les quelques cooccurrents de *Gleichung* communs aux deux corpus et dotés d'indices de spécificité élevés. Ils ne sont guère surprenants : outre plusieurs verbes usuels comme *genügen*, *befriedigen* (deux synonymes de « satisfaire ») ou *erhalten* (« obtenir »), on trouve le substantif *Elimination*, qui rappelle l'importance de ce procédé technique pour les mathématiciens de l'époque, ainsi que *Wurzel* (« racine »), un autre terme naturellement lié à *Gleichung* — l'indice de spécificité de la cooccurrence de ce dernier terme se trouve être toutefois relativement plus élevé du côté des invariants, ce qui est déjà un signe (qui sera confirmé dans quelques lignes) que le thème de la résolution des équations algébriques est y abordé en tant que tel.

La présence de *Form* dans la seule liste des cooccurrents de *Gleichung* du côté de la théorie des surfaces ne doit pas induire en erreur. Très majoritairement employé dans des expressions comme « une équation de la forme : # », il est en fait remplacé par son synonyme *Gestalt* en théorie des invariants [33]. Cette dissymétrie s'explique facilement : *Form*, qui désigne aussi les objets usuellement associés aux invariants que sont les formes algébriques (quadratiques binaires, par exemple) est évité dans les textes relatifs à ce sujet pour référer à l'aspect des équations ; réciproquement, *Gestalt* fait partie des mots relativement courants dans le corpus de la théorie des surfaces, mais est utilisé presque exclusivement pour décrire l'apparence extérieure de ces surfaces, et pas celle d'équations.

Mais d'autres lemmes cooccurrents à *Gleichung* sont propres (ou ont un indice de spécificité sans commune mesure) à l'un des deux corpus sans pour autant avoir de tels équivalents dans l'autre. Ainsi, les cooccurrents *homogen* et *Koordinate*, qui figurent uniquement du côté des surfaces, renvoient directement à la géométrie projective et ses équations entre coordonnées homogènes définissant divers objets, tandis que *Lösung*, *Auflösung* et *fünft*, que l'on trouve associés à *Gleichung* principalement en théorie des invariants, rappellent que la résolution algébrique des équations (en particulier du cinquième degré) fait partie des sujets abordés de front par plusieurs contributeurs au corpus des invariants (surtout Gordan, Brill et Klein). D'autres lemmes particuliers, au contraire, figurent parmi les cooccurrents les plus spécifiques de *Gleichung* par le fait d'un auteur seulement : ainsi, l'adjectif *determinierend* apparaît comme très fortement lié à *Gleichung* car David Hilbert définit puis utilise à de nombreuses reprises une notion technique d'équation de détermination (d'objets comme des formes ou des faisceaux de formes) dans un unique article de 1886 classé en théorie des invariants [Hilbert 1886].

Si ces cooccurrents-là reflètent donc des thématiques de recherche ou des objets impliqués de manière spécifique dans l'un ou l'autre des corpus, d'autres lemmes font voir d'autres caractéristiques du travail mathématique. Prenons l'exemple des trois lemmes *Seite*, *rechts* et *links*, qui se trouvent être en association forte avec *Gleichung* exclusivement dans les textes de théorie des invariants. Une inspection des cooccurrences situées dans leur contexte montre déjà que *Seite* fait presque toujours référence aux « membres » ou « côtés » d'une équation, précisés par les adjectifs *rechts* et *links* (« de droite », « de gauche ») — c'est seulement dans de rares cas que *Seite* est utilisé dans le sens de

33. Comme précédemment, toutes les explications et interprétations proposées ici s'appuient sur l'indispensable recours aux concordanciers et, plus largement, aux textes originaux complets. Le cooccurrent *Gestalt* n'apparaît pas dans la table 4 car il est situé plus bas dans le classement.

« page » ou de « côté » d'un polygone. En fait, l'emploi de ces termes est révélateur d'un type de travail technique que l'on rencontre beaucoup plus dans le corpus des invariants [34], à savoir que les différents membres des équations y sont régulièrement observés et soumis à diverses transformations ou manipulations algébriques avant d'être réinjectés dans d'autres équations ou d'être eux-mêmes égalés à zéro, ou encore de faire l'objet de commentaires descriptifs permettant aux auteurs d'avancer sur la question traitée. Or, si ces manières de faire semblent ne survenir dans le corpus des surfaces que bien plus rarement, c'est que beaucoup des équations centrées de ce corpus-là sont des équations qui définissent courbes, surfaces et autres lieux géométriques, et dont un des membres est systématiquement égal à zéro. De ce fait, elles forment un terrain moins propice aux mêmes pratiques que celles vues dans le corpus des invariants. Bien moins nombreuses que dans ce dernier, les équations centrées de la théorie des surfaces sont donc aussi sujettes à des pratiques différentes, dans la mesure où elles font moins l'objet de dissections, d'observations et de diverses transformations qui apparaissent comme plus capitales pour les invariants.

Plus généralement, que les équations, formules et autres identités sont davantage au cœur du travail mathématique dans le corpus des invariants se manifeste aussi à travers les verbes pleins (c'est-à-dire ni modaux, ni auxiliaires) employés de part et d'autre. Si l'on regarde dans un premier temps ceux de ces verbes qui sont le plus utilisés en nombres absolus (voir la table 5), on constate d'ores et déjà que le corpus des surfaces fait intervenir beaucoup de verbes exprimant ce que j'appellerai des « actions géométriques », comme *gehen*, *schneiden*, *liegen*, *entsprechen*, *berühren*, *treffen* (« passer », « couper », « être situé », « correspondre », « toucher », « rencontrer »), etc., dont les sujets et compléments associés sont presque toujours des objets comme des points, des courbes ou des surfaces [35]. À ces verbes d'action géométrique s'ajoutent quelques verbes plus fonctionnels pour le discours mathématique, comme *geben* ou *erhalten* (« donner » —

34. Il ne s'agit pas d'affirmer que ce type de travail n'existe pas du tout du côté des surfaces, mais qu'il survient de manière statistiquement anormalement élevée dans le corpus des invariants.

35. À première vue, un verbe comme *gehen* peut signifier bien d'autres choses (on le traduit généralement plus volontiers par « aller »), mais l'inspection des concordanciers et de ses cooccurrents (parmi lesquels *durch* est de loin le plus caractéristique) atteste de son emploi majoritaire en tant que verbe d'action géométrique, comme dans la phrase « *Die Kurve geht durch diese Punkte.* » : « La courbe passe par ces points. » De manière plus discrète, *gehen* apparaît aussi comme la racine du verbe à particule séparable [*in etwas*] *übergehen* (« se transformer [en quelque chose] »), qui a souvent pour sujets et compléments des objets géométriques mais aussi, quelquefois, des équations.

mais le fréquent « *es gibt* » se traduit par « il y a » —, « obtenir »). Ces derniers figurent aussi parmi les verbes pleins les plus utilisés dans le corpus des invariants, aux côtés d'autres verbes du même type, comme *folgen*, *bezeichnen*, *finden* (« suivre », « désigner », « trouver »). Les quelques autres verbes fréquents en théorie des invariants qui ont un sens plus technique (*verschwinden*, *bilden* : « s'annuler », « former ») sont pour la plupart également présents de manière non négligeable du côté des surfaces, alors que les verbes d'action géométrique sont très peu représentés en invariants.

Pour compléter cette image, le calcul des spécificités des verbes pleins permet de mettre en évidence ceux qui sont singulièrement présents ou absents d'un des corpus. Plus précisément, ce point de vue consiste à mesurer l'écart entre les fréquences observées et les fréquences qu'on attendrait théoriquement au sens suivant. Considérant le corpus des surfaces comme une partie particulière de l'ensemble des mots issus de la réunion des deux corpus, et supposant que les verbes pleins sont équidistribués au sein de cet ensemble total, on peut calculer, selon un modèle de loi hypergéométrique, le nombre de verbes pleins attendus dans une partie quelconque de taille égale à celle du corpus des surfaces. Il y a alors une spécificité (positive) significative, et donc une sur-représentativité, quand la fréquence observée dans le corpus des surfaces est très supérieure à la fréquence attendue, un indice de spécificité permettant de quantifier numériquement les écarts aux valeurs théoriques [36].

Conformément à ce qui a été décrit précédemment, et comme le montre la table 6, les verbes les plus spécifiques pour le corpus des surfaces sont les verbes d'actions géométriques déjà donnés. Au contraire, le classement par spécificités met en évidence, pour ce qui est des invariants, toute une série de verbes qui ne figurent pas parmi les plus fréquents et qui renvoient pour beaucoup directement au champ lexical du calcul et des manipulations algébriques (*ersetzen*, *multiplizieren*, *berechnen* : « remplacer », « multiplier », « calculer », etc.). En outre, on peut remarquer que les indices de spécificité sont bien plus élevés pour les surfaces que pour les invariants, ce qui signifie que les verbes les plus sur-représentés dans le corpus des invariants le sont tout de même moins que ceux du corpus des surfaces, c'est-à-dire que les verbes

36. Ces calculs permettent aussi de détecter les cas où les fréquences observées sont très inférieures aux fréquences attendues, et on parle alors de spécificité négative. Dans le cas où deux parties sont comparées, les spécificités positives dans une partie sont négatives dans l'autre, mais cette dualité cesse quand davantage de parties sont considérées. Je m'intéresserai ici surtout aux spécificités positives, omettant ainsi ce qualificatif dans mes descriptions.

Surfaces					
Verbe	F. Surf.		F. Inv.		τ
gehen	1 739	4,3 %	417	1,4 %	3,1
schneiden	1 659	4,1 %	99	0,3 %	13,7
liegen	1 594	3,9 %	219	0,7 %	5,6
geben	1 532	3,8 %	769	2,6 %	1,5
entsprechen	1 182	2,9 %	232	0,8 %	3,6
berühren	1 137	2,8 %	87	0,3 %	9,3
bilden	1 095	2,7 %	648	2,2 %	1,2
erhalten	1 033	2,6 %	1047	3,5 %	0,7
treffen	1 012	2,5 %	46	0,2 %	12,5
enthalten	888	2,2 %	645	2,2 %	1
lassen	872	2,2 %	951	3,2 %	0,7

Invariants					
Verbe	F. Inv.		F. Surf.		τ
erhalten	1 033	3,5 %	1 047	2,6 %	1,3
lassen	951	3,2 %	872	2,2 %	1,5
setzen	826	2,8 %	592	1,5 %	1,9
geben	769	2,6 %	1 532	3,8 %	0,7
ergeben	703	2,4 %	621	1,5 %	1,6
verschwinden	695	2,3 %	361	0,9 %	2,6
folgen	673	2,3 %	539	1,3 %	1,8
bilden	648	2,2 %	1 095	2,7 %	0,8
enthalten	645	2,2 %	888	2,2 %	1
bezeichnen	613	2,1 %	406	1,0 %	2,1
finden	482	1,6 %	595	1,5 %	1,1

Table 5 – Les lemmes des verbes pleins les plus employés, avec leurs fréquences absolues et relatives, au sein de chaque corpus. Il y a en tout 40 492 tels lemmes en théorie des surfaces, contre 29 577 en théorie des invariants. Les fréquences relatives sont calculées par rapport à ces dernières données. Le nombre τ est le quotient de la fréquence relative des surfaces par celle des invariants.

d'action géométrique marquent le corpus des surfaces davantage que ceux des manipulations algébriques pour les invariants.

Sans entrer autant dans les détails, ajoutons que l'étude des noms communs spécifiques à chacun de nos deux corpus confirme ces observations : au contraire des surfaces, les invariants comportent, outre les substantifs correspondant aux objets propres que sont les formes, les

Surfaces		Invariants	
Verbe	Spécificité	Verbe	Spécificité
schneiden	273,5	ersetzen	60,7
treffen	179,3	verschwinden	53,9
liegen	176,8	ausdrücken	52,6
berühren	171,0	multiplizieren	48,2
gehen	115,2	setzen	34,0
entsprechen	97,4	bezeichnen	30,5
abbilden	64,4	bedeuten	28,5
legen	29,5	berechnen	22,5
begegnen	28,6	auslassen	20,7
hindurchgehen	26,4	folgen	20,4
zerfallen	25,0	entstehen	20,3

Table 6 – Les lemmes des verbes pleins les plus spécifiques.

invariants et leurs diverses déclinaisons, des noms spécifiques comme *Faktor*, *Formel*, *Ausdruck* (« expression »), *Identität*, *Operation*, mais aussi *Gestalt*, qui souligne bien l'attention portée sur l'observation des formules.

Pour finir, il est également instructif de s'intéresser aux distributions grammaticales elles-mêmes au sein des corpus. La table 7 montre les premières catégories grammaticales en nombres absolus d'occurrences. Les noms communs (NN) sont les plus communs et sont suivis de la catégorie associée des articles (ART). Les symboles mathématiques [37] (XY), les prépositions (APPR), les adjectifs attributifs (ADJA), les virgules ($,), les points et signes de ponctuation doubles ($.) figurent aussi dans le haut du classement, bien qu'à des positions relatives différentes. Les adverbes (ADV) et les verbes pleins sous forme finie (VVFIN) clôturent, à la même position, ces classements partiels — rappelons que l'analyseur morpho-syntaxique Treetagger, utilisé par TXM, distingue tout 54 catégories grammaticales pour l'allemand.

Passer ensuite à l'examen des spécificités des parties du discours met en évidence quelques phénomènes intéressants (voir la table 8). Pour continuer sur la piste suivie dans cette section, on notera que le corpus des surfaces se caractérise (entre autres) par une sur-représentativité des noms communs, auxquels s'ajoutent les articles, les pronoms relatifs substitutifs (PRELS) ou attributifs (PRELAT), les pronoms attributifs

37. Cette catégorie grammaticale compte donc, dans notre cas, l'ensemble des symboles $*$ et #, qui ont d'ailleurs été choisis lors du formatage initial justement dans cette optique.

Surfaces			Invariants		
Catégorie	Fréquence	%	Catégorie	Fréquence	%
NN	113 692	18,0	NN	74 649	16,2
ART	72 469	11,4	ART	46 061	10,0
APPR	50 991	8,1	XY	41 080	8,9
XY	45 998	7,3	APPR	36 672	8,0
ADJA	39 134	6,2	$.	31 460	6,8
$,	38 649	6,1	ADJA	30 652	6,7
$.	37 886	6,0	$,	26 510	5,8
ADV	30 539	4,8	ADV	22 707	4,9
VVFIN	24 477	3,9	VVFIN	17 396	3,8

TABLE 7 – Les premières catégories grammaticales dans les deux corpus. Les pourcentages sont rapportés aux nombres totaux de mots.

indéfinis (PIAT) et les virgules, tandis que celui des invariants comporte notamment une part anormalement élevée de symboles mathématiques, de signes de ponctuation forts et de pronoms relatifs adverbiaux (PWAV) [38].

Surfaces		Invariants	
Catégorie	Spécificité	Catégorie	Spécificité
CARD	234,9	XY	216,7
ART	127,4	$.	71,3
NN	126,1	NE	50,2
PRELS	73,7	VVIZU	27,2
PIAT	50,9	PWAV	25,3
PRELAT	17,1	VMFIN	18,7
$,	13,8	PAV	17,2

TABLE 8 – Certaines des catégories grammaticales les plus spécifiques dans les deux corpus.

38. Pour les invariants, la sur-représentativité des catégories des noms propres (NE), des verbes pleins sous forme infinitive avec un *zu* (VVIZU, comme dans *auszudrücken*, qui exprime l'infinitif du verbe à particule séparable *ausdrücken* au sein de certaines propositions infinitives), des verbes modaux sous forme finie (VMFIN) et des adverbes pronominaux (PAV) me paraît délicate à interpréter à ce stade de l'enquête. En ce qui concerne les surfaces, la présence des nombres cardinaux (CARD, exprimés en toutes lettres ou en chiffres) semble pointer vers l'attention portée par les contributeurs au corpus sur les configurations formées d'un nombre fini objets géométriques divers. Je resterai toutefois prudent sur ce point, en particulier parce qu'une partie des cardinaux des textes originaux s'expriment par des formules mathématiques et sont donc comptés dans la catégorie XY.

Les indices récoltés dans cette section suggèrent finalement des modes d'écriture avec des tendances opposées : du côté des surfaces, des phrases amples, riches en propositions relatives [39], faisant apparaître les noms d'objets en toutes lettres (lesquels ne sont donc pas entièrement substitués par des symboles mathématiques [40]) et faisant intervenir des verbes pleins tout à fait spécifiques pour en décrire les interactions. Au contraire, des phrases plus ramassées du côté des invariants, davantage tournées vers des formules centrées avec leurs manipulations et calculs associés — l'immense majorité des pronoms relatifs adverbiaux, sur-représentés dans ce corpus algébrique, est concentrée dans l'emploi de *wo* (« où »), très souvent utilisé pour préciser les symboles qu'impliquent les formules centrées, comme dans : « On obtient : #, où $*$ désigne une fonction quelconque. »

Cette manière d'écrire particulière au corpus des surfaces, et qui émerge ainsi nettement par comparaison avec la théorie des invariants, n'est bien entendu pas une donnée absolue : ni en termes d'universalité en son sein — ce que nous avons décrit ici reste de l'ordre des tendances générales et ne dit rien des disparités pouvant exister entre les textes du corpus, dont certaines apparaîtront d'ailleurs dans la prochaine section —, ni, bien sûr, en termes de stabilité au cours du temps. Par exemple, il semblerait raisonnable de conjecturer que les réécritures de la géométrie algébrique qui surviennent au début du XX^e^ siècle tendent à faire fondre le nombre de ce que j'ai appelé les verbes d'action géométrique au profit des formules mathématiques, dont les symboles eux-mêmes se substitueraient à ces verbes.

4. Une analyse contrastive endogène

Le dernier volet de l'étude consiste à regrouper par proximité lexicale les textes du corpus des surfaces algébriques en plusieurs classes puis à analyser ces dernières en croisant différents points de vue.

De telles classes sont proposées par le logiciel TXM à l'issue d'un procédé de classification ascendante hiérarchique, dont le principe est le suivant. On commence par lancer sur l'ensemble du corpus une requête lexicale ou grammaticale : dans le cas présent, j'ai choisi de considérer tous les mots qui sont des noms (communs et propres), des adjectifs, des

39. En allemand, les propositions relatives qui suivent une proposition principale sont systématiquement introduites par des virgules.

40. L'alliance d'un désignant en toutes lettres et d'un symbole associé, comme dans « La surface F est singulière. » est toutefois monnaie courante. Je reparlerai brièvement de ce mode d'écriture plus bas.

adverbes, des verbes pleins ou des symboles mathématiques [41]. Chaque texte du corpus est alors assimilé à un vecteur dont les composantes sont égales aux fréquences, au sein du texte, des différents résultats de la requête trouvés dans le corpus entier. Une notion de distance étant définie entre de tels vecteurs-textes, un algorithme agrège, étape après étape, les deux textes (ou groupements de textes créés à l'étape précédente) les plus proches, jusqu'à n'obtenir plus qu'un groupement correspondant au corpus entier. Les traces conservées de ces appariements successifs (représentables sous la forme d'un dendrogramme : voir la figure 3) permettent finalement de partitionner le corpus en un certain nombre de classes dont les textes ont un profil lexical analogue. Ce nombre peut être choisi de manière arbitraire, mais des indicateurs numériques calculés au cours de l'algorithme permettent d'en apprécier la pertinence et donc d'en sélectionner une valeur optimale, au sens où les classes de la partition correspondante représentent au mieux les convergences et divergences lexicales au sein du corpus.

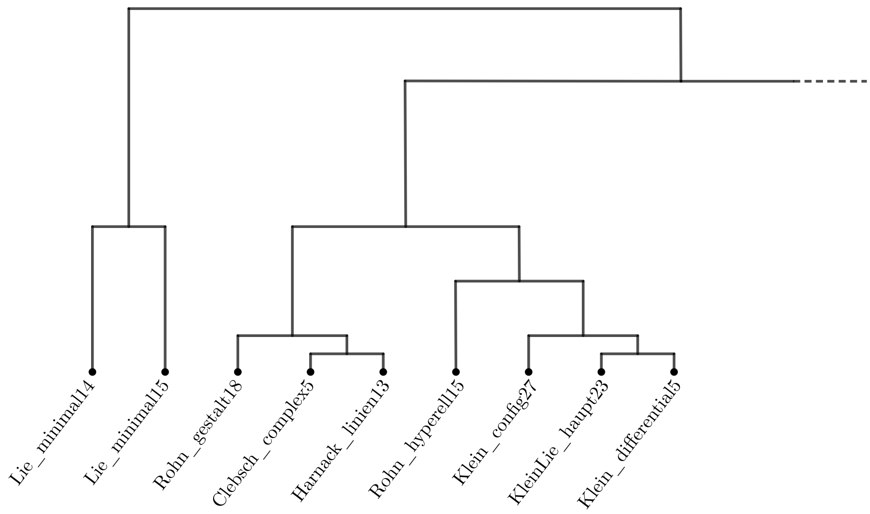

FIGURE 3 – Détail du dendrogramme représentant la classification ascendante hiérarchique. Chaque feuille représente un texte et chaque trait horizontal représente un appariement. Cet extrait représente les deux premières classes lexicales.

41. Considérer les résultats de cette requête en termes de mots ou de lemmes conduit à des classifications essentiellement identiques. Une telle stabilité a déjà été remarquée plusieurs fois dans le cas des textes littéraires : voir [BRUNET 2007], ainsi que [PINCEMIN 2020, § 2.4] et les références qui y sont citées.

Le corpus des surfaces algébriques a ainsi été scindé en six classes lexicales. Leur description sera faite en s'aidant des outils textométriques déjà présentés, mais aussi en les confrontant aux classifications disciplinaires offertes par le registre des *Mathematische Annalen* et par le *Jahrbuch über die Fortschritte der Mathematik*, ainsi qu'aux résultats obtenus par une analyse en termes de réseaux de citations [42]. En particulier, et c'est là un résultat qui me semble montrer à la fois l'utilité et la robustesse de l'approche textométrique dans ce cadre, les classes lexicales dégagées correspondent dans l'ensemble à autant de réseaux de citations.

La première classe a déjà été rencontrée : il s'agit des deux articles de Lie sur les surfaces minimales algébriques, [Lie 1879a,b]. Comme nous l'avons déjà remarqué, le vocabulaire qui y est employé comporte un grand nombre de spécificités qui renvoient à des objets et théorèmes liés à la géométrie différentielle, isolant ainsi ces deux textes du reste du corpus. Il en est de même au niveau des citations : à l'exception d'une référence bibliographique commune avec deux textes de la deuxième classe, aucune des autres références citées par Lie (principalement sur le sujet des surfaces minimales) n'est partagée par les autres textes du corpus. Cet isolement se retrouve également dans la classification des *Mathematische Annalen*, puisque ces articles de Lie sont les seuls du corpus à être aussi classés en géométrie différentielle — en revanche, un tel phénomène est inexistant dans le *Jahrbuch*, qui, de fait, ne propose pas de section consacrée à la géométrie différentielle à cette époque [43].

La deuxième classe reste de taille modeste, étant formée de sept articles publiés en parts équivalentes autour de 1872 d'abord, puis de 1880. Les listes des lemmes les plus fréquents et les plus spécifiques de cette classe montrent clairement que ces articles abordent plusieurs questions relatives à la théorie des complexes et des congruences de droites, et leur lien avec la surface de Kummer [44]. La présence tout aussi

42. Rappelons qu'il s'agit d'étudier les citations explicites et implicites effectuées dans les textes du corpus puis de regrouper ceux-ci en réseaux, dont les éléments se citent beaucoup entre eux tout en citant de manière plus rare (ou négative) ceux des autres réseaux, [Goldstein 1994, 1999]. L'étude de ces aspects s'est ici appuyée sur la base de données collective en ligne `Thamous` et ses différents outils, mis en place et entretenus par Alain Herreman.

43. Les sections de géométrie du *Jahrbuch* restent très stables au cours de la période qui est considérée ici, une grande réorganisation ne survenant qu'en 1916. Voir [Folta et Nový 1965, p. 16].

44. Complexes et congruences de droites sont des objets de la géométrie linéaire, ou *Liniengeometrie*, dans laquelle les objets de base considérés sont les droites de l'espace (et pas les points, par exemple). La surface de Kummer est une surface quartique possédant 16 points singuliers ordinaires, soit le maximum que peut avoir une telle

remarquable de lemmes comme *Periode*, *elliptisch*, *hyperelliptisch* ou *Theta* indique l'intervention répétée de la théorie des fonctions elliptiques et hyperelliptiques dans cette classe. Les deux principaux contributeurs sont ici Felix Klein et son élève Karl Rohn, qui se citent d'ailleurs beaucoup mutuellement, que ce soit pour des références du corpus ou non, comme dans le cas d'un texte de Klein de 1870 qui fait ici office de référence fondamentale, [Klein 1870]. Notons aussi qu'un court texte d'Axel Harnack [1878], un autre élève de Klein, abordant la question de certaines surfaces linéaires (*Linienflächen*) du quatrième ordre se trouve inclus dans cette classe lexicale bien qu'il soit un point isolé en termes de citations [45]. Tous les textes de cette deuxième classe relèvent des surfaces spéciales selon le *Generalregister* mais leur distribution est mixte pour le *Jahrbuch* : trois d'entre eux sont en effet classés en *Liniengeometrie*, deux sont en géométrie analytique de l'espace, un autre en théorie des fonctions et le dernier est à cheval entre ces deux dernières sections [46]. Il est ainsi intéressant de remarquer que la classification lexicale rapproche ici des textes issus de catégories disciplinaires distinctes.

C'est aussi le cas pour les quatorze textes de la troisième classe, publiés plutôt régulièrement entre 1871 et 1889 : du point de vue du *Jahrbuch*, ils sont pour la plupart sont éparpillés à travers différentes divisions de la géométrie analytique, mais quelques-uns se trouvent en géométrie synthétique et même en physique mathématique. Pour les *Mathematische Annalen*, en revanche, presque tous les textes traitent des surfaces spéciales. Les termes les plus spécifiques à cette classe décrivent le réseau lexical des singularités des surfaces : *Knoten* (« nœuds »), *Pinch-points* (un mot anglais repris tel quel), *Doppelkante* (« arête double »), *dreifach* (« triple »), *singulär* ou encore *isoliert*. Mais on trouve également d'autres mots qui font référence aux questions de détermination

surface. Les contributions de Klein et d'autres à ces sujets sont décrites dans [Rowe 1989, 2017].

45. Au contraire de ce qui est décrit dans [Goldstein 1999, p. 205], le corpus étudié ici, certes plus restreint, constitué différemment et se rapportant à un autre domaine mathématique, ne comporte que cinq tels textes isolés. Bien que nous n'y soyons pas présentement confrontés (sauf à considérer les deux articles de Lie de la première classe comme un seul texte), la classification lexicale peut tout à fait produire des classes réduites à un élément.

46. Le *Jahrbuch* comporte en fait une section entière consacrée à la géométrie analytique et divisée en plusieurs chapitres. L'un d'eux est intitulé « géométrie analytique de l'espace » (et est découpé en plusieurs sous-divisions) tandis qu'un autre regroupe la *Liniengeometrie*. Cette dernière fait donc partie de la géométrie analytique mais y est distinguée comme sujet. Ici comme dans la suite, lorsque je parlerai de la « géométrie analytique de l'espace », je ferai référence au chapitre correspondant, et pas à la section plus générale à laquelle il appartient.

de l'aspect des surfaces : *reell*, *imaginär*, *Mantel* (« nappe »), *Öffnung* (« ouverture »), *Ast* (« branche »), *Gestalt*, *Modell*, etc. Un examen plus approfondi des textes montre en fait que si tous s'intéressent d'une manière ou d'une autre à la question des singularités, environ un tiers d'eux n'évoque pas celle de la forme des surfaces — ceux qui le font cherchent parfois à élucider l'aspect local de celles-ci (autour des singularités) ou à en comprendre la forme globale, en s'aidant alors des points singuliers [47]. Au contraire de la deuxième classe lexicale, de nombreux types de surfaces sont au cœur des recherches, que ce soit des surfaces cubiques (générales, de Clebsch, à quatre points singuliers) ou quartiques (de Steiner, des ondes, avec un point triple, réglées). Rohn est ici l'auteur le plus prolifique, avec cinq textes sur les quatorze de la classe, suivi notamment de Klein et de Friedrich Emil Eckardt. Outre les inter-citations ou références communes fréquentes (parmi lesquelles on trouve par exemple un article de Ludwig Schläfli souvent reconnu comme de première importance pour la classification des surfaces cubiques, [Schläfli 1863]), les textes de cette classe lexicale entretiennent des liens directs un peu plus fournis vers la deuxième classe que vers les autres [48]. Par ailleurs, certains des articles sur les surfaces cubiques, [Gordan 1872 ; Eckardt 1872, 1876], partagent des références communes avec les quatrième, cinquième et sixième classes lexicales. Ces références sont toutefois souvent vues comme des travaux fondateurs ou de synthèse au sujet de certaines surfaces ; les citations qui en sont faites sont parfois de nature historique, et en tout cas rarement des emprunts techniques effectifs. Parmi ces références générales communes, on trouve par exemple [Sturm 1867 ; Cremona 1868], deux écrits de grande envergure présentant la théorie des surfaces cubiques en reprenant notamment toutes ses bases.

Les quatorze textes de la quatrième classe lexicale, publiés essentiellement entre 1872 et 1881, forment un réseau de citations clairement identifiable. Comme l'indique d'emblée le substantif le plus spécifique de cette classe, *Zahl* (« nombre »), les textes qui la composent abordent tous des questions énumératives. Cela est directement confirmé par le fait que tous les textes du corpus que le *Jahrbuch* place en géométrie énumérative se retrouvent ici. En outre, deux articles de la classe ne sont pas issus de cette division disciplinaire, mais abordent bien de telles questions : par exemple, le titre d'un article de Voss [1875] appartenant à cette classe

47. En ce qui concerne ce sujet, voir les références données en introduction à propos des modèles de surfaces.

48. De ce point de vue, il serait peut-être possible de parler d'un réseau formé des textes des deux classes, vis-à-vis du reste du corpus, et de deux sous-réseaux au sein de celui-ci.

montre qu'il a pour objectif de déterminer le nombre de points circulaires sur une surface générale (à ce propos, notons que presque tous les textes de la classe relèvent de la théorie générale des surfaces selon le *Generalregister*). D'autres termes de haute spécificité lexicale font référence à des objets ou personnes associés : *Koinzidenz* renvoie directement à l'utilisation répétée du principe de correspondance, tandis que les noms propres *Schubert* et *Zeuthen* sont les noms de deux acteurs essentiels des questions de géométrie énumérative de l'époque [49]. Il est également intéressant de relever la sur-représentativité du symbole # ainsi que des deux lemmes *Formel* et *Lösung*, qui indiquent qu'il s'agit d'une classe lexicale particulièrement riche en formules et autres équations centrées. Pour des raisons de langue, Zeuthen ne fait pas partie des contributeurs du corpus mais plusieurs de ses textes (effectivement publiés dans les *Mathematische Annalen*) forment un socle de références communes, auxquelles s'adjoignent deux célèbres textes de Hermann Schubert sur la géométrie énumérative, [Schubert 1876, 1879], souvent cités par les textes les plus tardifs de la classe. Schubert lui-même publie ici, mais l'auteur le plus prolifique est Aurel Voss, avec cinq articles. Rudolf Sturm, Heinrich Krey et Otto Rupp font également partie des auteurs participant à cette classe avec plus d'un texte, Sturm ayant la particularité d'être celui qui cite le plus les références générales mentionnées précédemment, communes à d'autres classes [50].

La cinquième classe lexicale voit Sturm être un de ses grands participants, avec cinq textes sur dix-sept, encore avec ce statut un peu particulier. Avant de préciser cela, notons qu'une première volée d'articles de la classe paraît entre 1869 et 1871, les autres étant ensuite publiés à peu près régulièrement entre 1878 et 1890. Presque tous ces articles sont situés dans une division ou l'autre de la « nouvelle géométrie synthétique » du *Jahrbuch*, alors que la distinction entre le général et le particulier faite par les *Mathematische Annalen* les sépare en deux parties de tailles presque égales. Theodor Reye, Friedrich Schur, Hermann Thieme et Ferdinand Affolter sont d'autres participants à cette classe lexicale. Une bonne partie des spécificités du vocabulaire, comme *Büschel*, *Bündel*, *erzeugen*, *Erzeugung* (« faisceau », « gerbe », « engendrer », « génération ») renvoie au thème de la génération d'objets géométriques à l'aide de familles d'ob-

49. Le sujet de l'émergence de la géométrie énumérative, surtout autour de la circulation de la théorie des caractéristiques de Chasles, a été récemment abordé dans la thèse de Nicolas Michel, [Michel 2020].

50. Il est intéressant de noter l'inclusion de travaux de Sturm dans cette quatrième classe car il est explicitement identifié par Schubert comme un mathématicien dont les travaux ne relèvent pas de la géométrie énumérative, pour des questions de méthode, [Michel 2020, p. 294].

jets inférieurs (typiquement, des surfaces à partir de familles de plans). Ce thème est transverse à la classe lexicale, qu'il apparaisse comme résultat intermédiaire ou comme enjeu en soi. Des aspects techniques importants qui lui sont liés sont évoqués par d'autres termes spécifiques, comme *assoziiert*, *projektiv*, *kollinear*, *Kollineation*. Si une bonne partie des introductions des articles de la classe mettent explicitement en avant qu'ils relèvent de la géométrie pure ou synthétique, et que savoir ainsi engendrer les surfaces est essentiel de ce point de vue disciplinaire [51], les objectifs visés restent relativement variés, et concernent différents types de surfaces (d'ordre donné, avec un nombre fini de droites, construites à partir d'autres surfaces, etc.). Pour ce qui est des citations, une sorte de dichotomie est à observer, comme dans la classe précédente. Si l'on excepte ceux de Sturm, les textes de la classe forment clairement un réseau de citations, avec des inter-citations nombreuses et des références communes, parmi lesquelles se trouvent notamment plusieurs textes de Reye, dont un des articles du corpus, [Reye 1869], ou le deuxième volume de la célèbre *Geometrie der Lage*, sous la forme de sa première ou de sa deuxième édition [Reye 1868, 1880]. En revanche, les textes de Sturm, qui ne citent pas les autres articles de la classe lexicale, ont plusieurs sources communes avec ceux-ci (y compris les textes de Reye évoqués à l'instant), et citent également certaines références du vivier commun aux autres classes. Pour en revenir aux données textométriques, remarquons enfin que cette cinquième classe lexicale est caractérisée par une sur-représentativité écrasante du symbole $*$ des formules mathématiques situées en plein texte et une relative sous-représentativité des noms communs et des formules centrées. Cette distribution semble renvoyer à une manière d'écrire les mathématiques qu'on retrouve fréquemment surtout dans ces textes, où, au cours de longs paragraphes sans aucune formule centrée, les noms d'objets exprimés dans le langage naturel sont très souvent suivis de, et parfois entièrement substitués par, leur nom spécifique incarné par une lettre ou une combinaison de lettres [52] (voir l'exemple donné en figure 4).

La sixième et dernière classe lexicale est la plus importante numériquement, avec vingt-deux textes. Contrairement aux autres classes, elle est très concentrée dans le temps, puisque que seuls trois de ses articles sont publiés après 1875. Ses contributeurs les plus prolifiques sont Clebsch ainsi que ses proches collègues Max Noether et Georg Korndörfer, avec

51. Voir par exemple [Reye 1869, p. 455] ou [Schur 1881, p. 1-2].

52. Formater et corriger les fichiers `txt` pour préparer le corpus permet déjà de sentir ce phénomène de manière très concrète, les interventions pour placer les symboles $*$ se démultipliant à l'extrême (et la fastidiosité de la tâche montant alors en flèche).

Haben wir nun umgekehrt auf einer G^2 drei Gerade g_1, g_2, g_3 und auf einer C^2 drei Gerade c_1, c_2, c_3 so, dass g_1 die c_2 und c_3, g_2 die c_3 und c_1, g_3 die c_1 und c_2 schneidet, so giebt es eine Fläche 2. Grades F^2, in Bezug auf welche g_1 und c_1, g_2 und c_2, g_3 und c_3 reciproke Polaren sind. Ordnen wir nämlich den Punkten (g_1, c_2), (g_1, c_3) die Ebenen $[c_1, g_2]$, $[c_1, g_3]$ und den Punkten (c_1, g_2), (c_1, g_3) die Ebenen $[g_1, c_2]$, $[g_1, c_3]$ als Polaren zu, so bilden alle dadurch bestimmten Flächen 2. Grades ein Büschel (oder eine Schaar) durch ein windschiefes Vierseit, dessen Ecken und Seitenflächen die Doppelelemente der durch jene Zuordnung auf g_1 und c_1 bestimmten Involutionen sind. Dem Punkte (g_2, c_3) entsprechen daher in Bezug auf

FIGURE 4 – Extrait (p. 11) d'un article de Schur appartenant à la cinquième classe lexicale, [SCHUR 1881].

respectivement six, cinq et quatre articles. Notons que Noether publie ici ses deux articles toujours retenus dans le récit historique usuel de la théorie des surfaces algébriques, [NOETHER 1870, 1875], mais d'autres textes de lui font aussi partie de la classe et y sont d'ailleurs davantage cités, comme un long mémoire sur les surfaces possédant des familles de courbes rationnelles, [NOETHER 1871]. Plus globalement, la plupart des textes de la classe concernent les surfaces spéciales d'après le *Generalregister*[53] et appartiennent au chapitre du *Jahrbuch* dévolu aux « correspondances, transformations univoques et représentations » et inclus dans la section de géométrie analytique. Plus précisément, c'est la thématique des représentations de surfaces que les spécificités lexicales les plus hautes font ressortir : *Abbildung* et *abbilden* (« représentation », « représenter ») sont très haut dans le classement et sont complétés par d'autres termes comme *Fundamentalpunkt*, qui fait référence à des objets centraux de ce problème des représentations[54]. D'autres termes à forte spécificité sont en tant que tels moins directement évocateurs du thème de la représentation des surfaces, mais renvoient en fait à des problèmes usuels qui y sont associés. Par exemple, *Kurve* et *Ordnung* sont deux des lemmes les plus spécifiques de cette classe lexicale bien qu'étant des mots tout à fait courants en géométrie — on les trouve un peu partout parmi les hautes fréquences absolues. Leur suremploi dans cette classe

53. Les deux fameux textes de Noether [1870, 1875] font partie des quelques exceptions, ce qui n'est peut-être pas étranger à leur place de choix dans l'historiographie.

54. En termes actuels, il s'agit d'établir une transformation birationnelle entre une surface et une autre, qui s'avère souvent être le plan projectif dans les articles dont il est question ici. Les points fondamentaux sont alors des points du plan qui correspondent non pas à des points, mais à des courbes tracées sur la surface de départ.

lexicale trahit une question de première importance dans la théorie des représentations de surfaces : comprendre comment les courbes tracées sur une surface se transforment par la représentation, notamment au niveau de leur ordre. Pour ce qui est des parties du discours, en comparaison avec les autres classes lexicales, celle-ci contient une part anormalement élevée de noms communs, d'articles et de pronoms relatifs. Ainsi, ce que nous avons décrit sur les emplois non équidistribués de phrases longues entre théorie des surfaces et théorie des invariants se retrouve d'une certaine façon ici, bien qu'à une échelle d'observation et de comparaison plus restreinte. Avant de conclure, notons encore que les textes de cette classe sont marqués par un très fort degré d'inter-citations. Ils partagent aussi de nombreuses références communes, parmi lesquelles les plus universellement citées sont des textes de Clebsch lui-même, dont son célèbre mémoire sur l'application des fonctions abéliennes à la géométrie, [Clebsch 1864], mais aussi un article, le plus cité de tous, sur le sujet des représentations de surfaces, en particulier des quatrième et cinquième ordres, [Clebsch 1869]. De ce point de vue, ce dernier apparaît donc comme plus central au sein de la classe lexicale que l'article de Clebsch, lui aussi cité mais de façon moindre, présentant pour la première fois la représentation des surfaces du troisième ordre, bien mieux connu dans l'histoire de la théorie des surfaces algébriques, [Clebsch 1866].

* *
*

Il est clair que les six classes lexicales pourraient être étudiées de façon plus fouillée, y compris du point de vue textométrique, en y examinant par exemple la distribution des noms propres, les spécificités du vocabulaire (voire de la grammaire) de certains auteurs ou groupes d'auteurs, ou encore les diverses colorations sémantiques de termes choisis. Les possibilités offertes par l'analyse factorielle des correspondances, technique complémentaire à la classification ascendante hiérarchique qui a été passée sous silence ici, pourraient elles aussi être explorées dans cette optique[55].

Bien qu'encore préliminaire de ce point de vue, l'étude présentée dans cette section met toutefois en évidence plusieurs résultats qui me paraissent intéressants et méritent quelques commentaires.

La structuration du corpus découlant de la classification lexicale, doublée de l'étude des termes spécifiques aux classes résultantes, a ainsi

55. Des exemples d'utilisation de cette technique sont présentés dans les références générales sur la textométrie données en introduction.

mis en lumière certaines thématiques de recherche peu visibles dans l'historiographie du sujet, comme l'étude des surfaces du point de vue de la « nouvelle géométrie synthétique » ou de celui des représentations de surfaces spéciales, en en montrant notamment les termes spécifiques et leur signification. Étudier ensuite, sans nécessairement faire appel aux techniques textométriques elles-mêmes, les contributeurs des différentes classes et les liens entre leurs textes a aussi mis en valeur des textes clés, comme [Clebsch 1869], tout autant que quelques mathématiciens dont le détail des travaux en théorie des surfaces est encore relativement peu connu : c'est par exemple le cas de Friedrich Schur, participant à la nouvelle géométrie synthétique, d'Aurel Voss, qui apparaît comme un contributeur tout à fait significatif en géométrie énumérative, mais aussi de Rudolf Sturm, dont la présence dans différentes classes lexicales et la position d'interface entre celles-ci intriguent et appellent, à l'instar des autres pistes dévoilées, à une investigation plus poussée.

Sur un plan plus méthodologique, la confrontation effectuée entre trois types de classifications du corpus (correspondant à l'approche du lexique, des citations ou des rubriques disciplinaires de l'époque) est également instructive. J'ai déjà eu l'occasion de souligner que les classes lexicales se superposent bien aux réseaux de citations que l'on peut dégager. Il est donc remarquable que de tels réseaux se retrouvent par l'inspection seule du lexique, ou, plus précisément, d'une partie du lexique, puisque la classification s'est basée seulement sur les catégories grammaticales pleines et les nombres de symboles tenant lieu des formules mathématiques. Le vocabulaire des textes mathématiques permet donc de retrouver les dynamiques historiques que reflètent les citations [56], ce qui suggère réciproquement que la participation d'un mathématicien à un groupe de travaux correspondant à un réseau de citations s'accompagne d'une certaine contrainte (au sens social) à adopter le lexique qui y est collectivement employé — le cas de Sturm, qui publie dans plusieurs classes lexicales est exemplaire à ce titre.

Notons d'ailleurs que la classification lexicale ne semble pas retrouver directement les réseaux de personnes, sauf si celles-ci se manifestent dans les sujets mathématiques eux-mêmes. Regardons par exemple la sixième classe lexicale, concernant le sujet des représentations de surfaces. Comme indiqué plus haut, ses contributeurs principaux sont Clebsch et deux

56. Tant les classes lexicales que les réseaux de citations ont été dégagés sans référence à des textes extérieurs au corpus : il est donc impossible en l'état d'évaluer dans quelle mesure ces regroupements se retrouveraient tels quels au sein de classes et réseaux plus vastes, constitués à partir d'un corpus plus étendu. Les dynamiques historiques dont il est question sont donc bien relatives au corpus considéré ici.

de ses proches, Noether et Korndörfer. Mais on y trouve aussi le texte correspondant à la thèse de Josef Diekmann [1871], la première que Klein a dirigée [57] et justement consacrée au sujet des représentations, alors que ce dernier, proche élève de Clebsch vers 1870, publie essentiellement dans les deuxième et troisième classes lexicales. Deux autres situations ont été vues précédemment : celle de Harnack, étudiant de Klein, dont un texte est isolé au point de vue des citations mais rattaché par le lexique à la deuxième classe ; celle de Voss, proche de Clebsch et de Klein, qui publie uniquement dans la classe de géométrie énumérative, dont est pourtant absent ce dernier. D'autres cas analogues peuvent être observés ; ils témoignent de ce que les réseaux de personnes et les filiations scientifiques ne se superposent que partiellement aux classes lexicales.

Revenons enfin sur la manière dont le vocabulaire agit vis-à-vis des rubriques disciplinaires du *Jahrbuch* : trois des classes lexicales peuvent être caractérisées, dans l'ensemble, par l'appartenance respective de leurs textes à l'une de ces rubriques (les questions énumératives, la nouvelle géométrie synthétique et les représentations de surfaces), tandis que d'autres relèvent de mixtes de différentes sections. La distinction entre surfaces générales et particulières proposée par les *Mathematische Annalen* se retrouve quant à elle très incomplètement à travers les classes lexicales [58]. La classification lexicale semble donc avoir tendance à ne respecter que partiellement les classifications disciplinaires du XIXᵉ siècle, rapprochant en particulier des textes situés dans différentes sections. De plus, comme on l'a vu, ce phénomène est susceptible de prendre différentes formes : rattachement de quelques textes à une thématique par ailleurs bien identifiée (c'est le cas des deux articles de Voss regroupés avec le reste de la géométrie énumérative) ou réorganisation plus large, comme pour les deuxième et troisième classe, de textes issus de diverses rubriques.

Toutes ces observations n'ont bien sûr pour le moment qu'une valeur strictement locale, et il serait intéressant de tester si les points de concours et les écarts relevés entre les différentes classifications se retrouvent dans d'autres situations historiques, avec des corpus peut-être constitués différemment : on pourrait par exemple se demander si, partant de l'intégralité des articles listés par le *Jahrbuch* dans les trois sections de géométrie énumérative, de nouvelle géométrie synthétique et de représentations des surfaces, la classification lexicale redonnerait trois

57. Voir [Tobies 2019, p. 95].

58. Demander une classification lexicale en deux parties ne permet pas de restituer la séparation entre le général et le particulier.

classes correspondantes.

5. Formules, polysémie et synonymie

L'enquête textométrique proposée repose sur plusieurs hypothèses de travail, dont certaines ont déjà été explicitées et qui ont pesé à la fois sur le façonnement du questionnaire auquel a été soumis le corpus et sur l'interprétation des données issues des calculs statistiques. Je voudrais revenir sur trois d'entre elles pour clore cet article.

D'abord, il est clair que l'opération qui a le plus radicalement contribué à transformer le corpus lors de sa préparation technique est l'arasement total du contenu des formules mathématiques. Comme expliqué plus haut, tenir compte de tels contenus dans toute leur complexité sémantique et syntaxique me semble être une tâche encore hors de portée pour l'instant, d'autant plus que les logiciels disponibles ne sont pas adaptés à ce type de question. En attendant que des solutions techniques adéquates voient peut-être le jour, il faut noter que si la perte d'information correspondante est indéniable pour le traitement textométrique en lui-même, il n'est pas non plus question d'affirmer que le contenu des formules doit rester irrévocablement invisible au cours de l'étude. Au contraire, l'indispensable va-et-vient entre les données statistiques et les textes originaux invite naturellement à le considérer dans toute son étendue afin de proposer une interprétation historique la mieux informée possible.

Ce n'est pas non plus dire que l'intégralité de l'information d'un texte mathématique s'incarne dans ses formules. Les résultats présentés montrent en effet que prendre en compte le vocabulaire associé à *Gleichung* ou même simplement le décompte des formules, centrées ou en plein texte, permet de mettre en évidence quelques faits intéressants relatifs à ces dernières. De plus, la distribution du corpus en six classes lexicales et sa cohérence générale avec le point de vue des citations suggèrent même que le lexique mathématique incarné par les mots du langage naturel pourrait d'une certaine manière saturer les textes originaux, et serait donc suffisant pour saisir dans l'ensemble de tels phénomènes collectifs. Autrement dit, il y aurait une redondance d'information entre les mots et les formules suffisamment forte pour que ces dernières ne soient pas indispensables à l'obtention de classifications pertinentes de textes.

Une autre hypothèse de travail a été utilisée de manière plus subreptice : celle de la stabilité sémantique du lexique des textes considérés vis-à-vis du procédé de classification lexicale. Par exemple, le bien-fondé des regroupements en classes que nous avons décrits suppose implicitement que la présence de mêmes mots dans différents textes est un indice de proximité lexicale. En un sens, c'est donc négliger, dans ce procédé, les cas de polysémie tant synchroniques que diachroniques : tout comme il existe des mots revêtant plusieurs significations dans des textes publiés simultanément (l'exemple de *Form* a été rencontré plus haut), les éventuels glissements de sens de certains termes (techniques ou non) au cours du temps sont aussi à envisager. Dans le cas soumis ici à l'étude, la cohérence des résultats obtenus et les différentes explorations du corpus qui ont été faites par ailleurs semblent indiquer que de tels problèmes de polysémie sont rares, ou du moins qu'ils ont un poids statistique suffisamment faible par rapport à la totalité des mots du corpus pour ne pas induire des rapprochements textuels inadéquats. Reste qu'il semble nécessaire de tenir compte de ce problème en puissance pour l'étude d'autres cas, notamment pour des investigations portant sur des périodes temporelles plus larges.

On pourrait enfin se demander si un phénomène inverse, de synonymie cette fois, pourrait survenir et ainsi perturber l'interprétation des résultats de classification lexicale : dans le cas où des termes distincts auraient des sens identiques, cette classification aurait tendance à séparer des textes qui mériteraient au contraire d'être rassemblés. Ce serait, je crois, occulter deux aspects. D'abord, encore une fois, c'est que les classifications issues de la textométrie sont aptes à agir en prenant le lexique dans toute sa masse, et donc à neutraliser ces éventuels cas problématiques dans la mesure où ils restent numériquement peu importants. Par ailleurs, en oubliant même l'enjeu classificatoire, savoir détecter les cas de synonymie est peut-être moins anodin que ce qu'il n'y paraît, le danger d'anachronisme étant susceptible de se dissimuler derrière cet acte. Reconnaître en effet un peu rapidement que deux mots sont des synonymes peut avoir pour conséquence de gommer des différences de perception essentielles existant aux yeux des mathématiciens du passé. Par exemple, j'ai montré ailleurs que l'existence de deux mots distincts, « genre » et « connectivité », aujourd'hui souvent compris comme des synonymes, reflétait de véritables différences de points de vue sur des objets et des disciplines associées, différences se traduisant elles-mêmes en des dynamiques historiques propres, [Lê 2020]. Ainsi, si les synonymies existent dans le lexique spécialisé des mathématiques, il me semble qu'elles aussi ont une histoire, qu'il convient de comprendre avant d'en tirer d'éventuelles conclusions pour une approche textométrique.

Ces quelques réflexions et mises en garde méthodologiques ne sont qu'un échantillon de celles susceptibles de naître d'un indispensable questionnement sur l'utilisation de l'outillage textométrique en histoire des mathématiques. Elles ne sont ni à balayer d'un revers de main, ni à être considérées comme disqualifiant cet outillage sans autre forme de procès. Ainsi, si elle a permis d'aborder d'une nouvelle manière une partie de l'histoire de la géométrie algébrique, la textométrie aura aussi été mise réciproquement à l'épreuve de la théorie des surfaces tout au long de cet article : comme pour d'autres méthodes, il s'agit bien de ne pas lui accorder une foi aveugle et de garder sur elle un regard toujours critique afin d'en contrôler les résultats, mais aussi, par effet de contraste, d'en apprécier toutes les possibilités.

Remerciements. — Je souhaite dédier cet article à Catherine Goldstein, que je remercie par ailleurs très chaleureusement pour ses remarques et ses suggestions. Je remercie également les rapporteurs anonymes et les éditeurs de la revue pour leurs retours sur le manuscrit.

Références

Affolter Ferdinand (1887), « Ueber Gruppen gerader Linien auf Flächen höherer Ordnung. (Zweite Mittheilung) », *Mathematische Annalen* **29**, p. 1-26 (↑ 9).

Barrow-Green June (2020), « "Knowledge gained by experience": Olaus Henrici—Engineer, Geometer and Maker of Mathematical Models », *Historia Mathematica* **54**, p. 41-76 (↑ 4).

Brigaglia Aldo (2016), « Picard and the Italian Mathematicians: The History of Three *Prix Bordin* », in Frédéric Brechenmacher, Guillaume Jouve, Laurent Mazliak et Rossana Tazzioli (éd.), *Images of Italian Mathematics in France: The Latin Sisters, from Risorgimento to Fascism*, Birkhäuser, p. 93-126 (↑ 4).

Brigaglia Aldo, Ciliberto Ciro et Pedrini Claudio (2004), « The Italian School of Algebraic Geometry and Abel's Legacy », in Olav Arnfinn Laudal et Ragni Piene (éd.), *The Legacy of Niels Henrick Abel*, Berlin, Heidelberg, New York : Springer, p. 295-347 (↑ 4).

Brunet Étienne (2007), « Le corpus conçu comme une boule », in François Rastier et Michel Ballabriga (éd.), *Corpus en Lettres et Sciences sociales : des documents numériques à l'interprétation. Actes du XXVII^e colloque d'Albi,* Langages et signification, textes publiés par Carine Duteil-Mougel et Baptiste Foulquié, Toulouse : Presses universitaires de Toulouse, p. 69-78 (↑ 26).

— (2016), *Tous comptes faits : écrits choisis. Tome III : Questions linguistiques*, Bénédicte Pincemin (éd.), Paris : Honoré Champion (↑ 2).

Casnati Gianfranco, Conte Alberto, Gatto Letterio, Giacardi Livia, Marchisio Marina et Verra Alessandro (éd.) (2016), *From Classical to Modern Algebraic Geometry: Corrado Segre's Mastership and Legacy*, Basel : Birkhäuser (↑ 4).

Clebsch Alfred (1866), « Die Geometrie auf den Flächen dritter Ordnung », *Journal für die reine und angewandte Mathematik* **65**, p. 359-380 (↑ 33).

— (1869), « Ueber die Abbildung algebraischer Flächen, insbesondere der vierten und fünften Ordnung », *Mathematische Annalen* **1**, p. 253-316 (↑ 5, 33, 34).

— (1864), « Ueber die Anwendung der Abelschen Functionen in der Geometrie », *Journal für die reine und angewandte Mathematik* **63**, p. 189-243 (↑ 33).

Cléry Matthias (2020), « La théorie des probabilités et l'Institut Henri Poincaré (1918-1940) : construction d'un champ probabiliste et pratique d'un transfert culturel », thèse de doct., Université Paris-Saclay (↑ 2).

Cremona Luigi (1868), « Mémoire de géométrie pure sur les surfaces du troisième ordre », *Journal für die reine und angewandte Mathematik* **68**, p. 1-133 (↑ 29).

Diekmann Josef (1871), « Ueber die Modificationen, welche die ebene Abbildung einer Fläche 3$^{\text{ter}}$ Ordnung durch Auftreten von Singularitäten erhält », *Mathematische Annalen* **4**, p. 442-475 (↑ 35).

Dieudonné Jean (1974), *Cours de géométrie algébrique*, t. 1. Aperçu historique sur le développement de la géométrie algébrique, Vendôme : Presses Universitaires de France (↑ 4).

Duran Samson (2019), « Des géométries étatsuniennes à partir de l'étude de l'*American Mathematical Society* : 1888-1920 », thèse de doct., Université Paris-Sud (↑ 2, 3).

Eckardt Friedrich Emil (1872), « Beiträge zur analytischen Geometrie des Raumes, insbesondere zur Theorie der Flächen 3$^{\text{ten}}$ Grades mit 4 Doppelpunkten und der Steiner'schen Flächen, sowie zur Lehre von den Raumcurven », *Mathematische Annalen* **5**, p. 30-49 (↑ 29).

— (1876), « Ueber diejenigen Flächen dritten Grades, auf denen sich drei gerade Linien in einem Punkte schneiden. (Aus einem Programm der Realschule I. Ordnung zu Chemnitz) », *Mathematische Annalen* **10**, p. 227-272 (↑ 29).

Fattori Marta (1980), *Lessico del* Novum organum *di Francesco Bacone*, Rome : Edizioni dell'Ateneo e Bizzarri Roma. Deux volumes (↑ 2).

FOLTA Jaroslav et NOVÝ Luboš (1965), « Sur la question des méthodes quantitatives dans l'histoire des mathématiques », *Acta historiae rerum naturalium nec non technicarum* **1**, p. 3-35 (↑ 27).

GOLDSTEIN Catherine (1994), « La théorie des nombres dans les Notes aux comptes rendus de l'Académie des sciences (1870-1914) : un premier examen », *Rivista di Storia della scienza*, 2e sér. **2** (2), p. 137-160 (↑ 27).

— (1999), « Sur la question des méthodes quantitatives en histoire des mathématiques : le cas de la théorie des nombres en France (1870-1914) », *Acta historiae rerum naturalium nec non technicarum* **3**, p. 187-214 (↑ 27, 28).

— (2012), « Les autres de l'un : deux enquêtes prosopographiques sur Charles Hermite », in Philippe NABONNAND et Laurent ROLLET (éd.), *Les uns et les autres... Biographies et prosopographies en histoire des sciences*, Nancy : Presses Universitaires de Nancy, p. 509-540 (↑ 12).

— (2018), « Hermite and Lipschitz: A Correspondence and Its Echoes », in Maria Teresa BORGATO, Erwin NEUENSCHWANDER et Irène PASSERON (éd.), *Mathematical Correspondences and Critical Editions*, Basel : Birkhäuser, p. 167-193 (↑ 7).

GORDAN Paul (1872), « Ueber das Pentaeder der Flächen dritter Ordnung », *Mathematische Annalen* **5**, p. 341-377 (↑ 29).

GRAY Jeremy (1989), « Algebraic Geometry in the Late Nineteenth Century », in John MCCLEARY et David E. ROWE (éd.), *The History of Modern Mathematics*, t. 1. Ideas and their Reception, Boston, San Diego, New York : Academic Press, p. 361-385 (↑ 4).

GRAY Jeremy, HASHAGEN Ulf, HOFF KJELDSEN Tinne et ROWE David E. (éd.) (2015), « History of Mathematics: Models and Visualization in the Mathematical and Physical Sciences », *Oberwolfach Reports* **47/2015** (↑ 4).

GUILBAUD Alexandre, PASSERON Irène, BARRELLON Vincent et FERRET Olivier (2014), « Éditer l'*Encyclopédie* au 21e siècle : un projet d'édition numérique, critique et collaborative », *Dix-huitième siècle* **46** (1), p. 153-166 (↑ 7).

HARNACK Axel (1878), « Bemerkungen zur Geometrie auf den Linienflächen vierter Ordnung », *Mathematische Annalen* **13**, p. 49-52 (↑ 28).

HEIDEN Serge, MAGUÉ Jean-Philippe et PINCEMIN Bénédicte (2010), « TXM : Une plateforme logicielle open-source pour la textométrie — conception et développement », in, *Proceedings of the 10th International Conference on the Statistical Analysis of Textual Data*, t. 2, Rome : Edizioni Universitarie di Lettere Economia Diritto, p. 1021-1032 (↑ 6).

HILBERT David (1886), « Ueber einen allgemeinen Gesichtspunkt für invariantentheoretische Untersuchungen im binären Formengebiete », *Mathematische Annalen* **28**, p. 381-446 (↑ 19).

— (1899/2015), *Grundlagen der Geometrie (Festschrift 1899)*, Klaus VOLKERT (éd.), Berlin, Heidelberg : Springer (↑ 17).

HOUZEL Christian (2002), *La géométrie algébrique : recherches historiques*, Paris : Albert Blanchard (↑ 4).

ISRAEL Giorgio (éd.) (2017), *Correspondence of Luigi Cremona (1830–1903)*, t. 1, Turnhout : Brepols (↑ 14).

JACQUES Marie-Paule et TUTIN Agnès (éd.) (2018), *Lexique transversal et formules discursives en sciences humaines*, Londres : ISTE Éditions (↑ 2).

JURISH Bryan (2011), « Finite-state Canonicalization Techniques for Historical German », thèse de doct., Universität Potsdam (↑ 8).

KLEIN Felix (1870), « Zur Theorie der Linienkomplexe des ersten und zweiten Grades », *Mathematische Annalen* **2**, p. 198-226 (↑ 28).

KOLMOGOROV Andrei Nikolaievitch et YUSHKEVICH Adolph-Andrei Pavlovich (éd.) (1981/1996), *Mathematics of the 19th Century*, Basel, Boston, Berlin : Birkhäuser. Traduit du russe (1981) par Roger Cooke (↑ 4).

LÊ François (2015), « "Geometrical Equations": Forgotten Premises of Felix Klein's *Erlanger Programm* », *Historia Mathematica* **42** (3), p. 315-342 (↑ 4).

— (2020), « "Are the *genre* and the *Geschlecht* one and the same number?" An inquiry into Alfred Clebsch's *Geschlecht* », *Historia Mathematica* **53**, p. 71-107 (↑ 37).

LEBART Ludovic, PINCEMIN Bénédicte et POUDAT Céline (2019), *Analyse des données textuelles*, Presses de l'Université du Québec (↑ 2).

LEMERCIER Claire et ZALC Claire (2008), *Méthodes quantitatives pour l'historien*, Paris : La Découverte (↑ 1).

LIE Sophus (1879a), « Beiträge zur Theorie der Minimalflächen. I : Projectivische Untersuchungen über algebraische Minimalflächen », *Mathematische Annalen* **14**, p. 331-416 (↑ 15, 27).

— (1879b), « Beiträge zur Theorie der Minimalflächen. II : Metrische Untersuchungen über algebraische Minimalflächen », *Mathematische Annalen* **15**, p. 465-506 (↑ 15, 27).

MCCLEARY John et ROWE David E. (éd.) (1989), *The History of Modern Mathematics*, t. 1. Ideas and their Reception, Boston, San Diego, New York : Academic Press.

MICHEL Nicolas (2020), « Of Words and Numbers: The writing of generality in the emergence of enumerative geometry (1852–1893) », thèse de doct., Université de Paris (↑ 30).

NOETHER Max (1870), « Zur Theorie des eindeutigen Entsprechens algebraischer Gebilde von beliebig vielen Dimensionen », *Mathematische Annalen* **2**, p. 293-316 (↑ 32).

— (1871), « Ueber Flächen, welche Schaaren rationaler Curven besitzen », *Mathematische Annalen* **3**, p. 161-277 (↑ 32).

— (1875), « Zur Theorie des eindeutigen Entsprechens algebraischer Gebilde. Zweiter Aufsatz », *Mathematische Annalen* **8**, p. 495-533 (↑ 32).

PINCEMIN Bénédicte (2020), « La textométrie en question », *Le Français moderne. Revue de linguistique française* **88** (1), p. 26-43 (↑ 2, 26).

PINCEMIN Bénédicte et HEIDEN Serge (2008), *Qu'est-ce que la textométrie ? Présentation*, http://textometrie.ens-lyon.fr/spip.php?rubrique80 (visité le 16/12/2020) (↑ 1).

REYE Theodor (1868), *Die Geometrie der Lage*, 1re éd., t. 2, Hannover : Carl Rümpler (↑ 31).

— (1869), « Projectivische Erzeugung der allgemeinen Flächen dritter, vierter und beliebiger Ordnung durch Flächenbündel niederer Ordnung », *Mathematische Annalen* **1**, p. 455-466 (↑ 31).

— (1880), *Die Geometrie der Lage*, 2e éd., t. 2, Hannover : Carl Rümpler (↑ 31).

ROWE David E. (1989), « The Early Geometrical Works of Sophus Lie and Felix Klein », in John MCCLEARY et David E. ROWE (éd.), *The History of Modern Mathematics*, t. 1. Ideas and their Reception, Boston, San Diego, New York : Academic Press, p. 209-273 (↑ 28).

— (2017), « Segre, Klein, and the Theory of Quadratic Line Complexes », in Gianfranco CASNATI, Alberto CONTE, Letterio GATTO, Livia GIACARDI, Marina MARCHISIO et Alessandro VERRA (éd.), *From Classical to Modern Algebraic Geometry: Corrado Segre's Mastership and Legacy*, Basel : Birkhäuser, p. 243-263 (↑ 4, 28).

ROWE David E. et TOBIES Renate (éd.) (1990), *Korrespondenz Felix Klein – Adolph Mayer*, Leipzig : Teubner (↑ 3).

SCHAPPACHER Norbert (2015), « Remarks about Intuition in Italian Algebraic Geometry », *Oberwolfach Reports* **47/2015**, p. 2805-2807 (↑ 4).

SCHLÄFLI Ludwig (1863), « On the Distribution of Surfaces of the Third Order into Species, in Reference to the Absence or Presence of Singular Points, and the Reality of Their Lines », *Philosophical Transactions of the Royal Society of London* **153**, p. 193-241 (↑ 29).

Schubert Hermann (1876), « Beiträge zur abzählenden Geometrie. Erste Abhandlung », *Mathematische Annalen* **10**, p. 1-116 (↑ 30).

— (1879), *Kalkül der abzählenden Geometrie*, Leipzig : Teubner (↑ 30).

Schur Friedrich (1881), « Ueber die durch collineare Grundgebilde erzeugten Curven und Flächen », *Mathematische Annalen* **18**, p. 1-32 (↑ 31, 32).

Sturm Rudolf (1867), *Synthetische Untersuchungen über Flächen dritter Ordnung*, Leipzig : Teubner (↑ 29).

Teissier Pierre, Quantin Matthieu et Hervy Benjamin (2018), « Humanités numériques et archives orales : cartographies d'une mémoire collective sur les matériaux », *Cahiers François Viète*, 3$^{\text{e}}$ sér. **4**, p. 141-177 (↑ 2).

Tobies Renate (2019), *Felix Klein: Visionen für Mathematik, Anwendungen und Unterricht*, Berlin, Heidelberg : Springer (↑ 17, 35).

Tutin Agnès (éd.) (2007), « Lexique des écrits scientifiques », *Revue française de linguistique appliquée* **12** (↑ 2).

Volkert Klaus (2019), « Note on Models », *Historia Mathematica* **48**, p. 87-95 (↑ 4).

Voss Aurel (1875), « Ueber die Zahl der Kreispunkte einer allgemeinen Fläche n^{ter} Ordnung », *Mathematische Annalen* **9**, p. 241-244 (↑ 29).