HAL
open science

Enjeux de la classification pour les mesures et les représentations de la compréhensibilité de textes

Adrien Barbaresi

To cite this version:

Adrien Barbaresi. Enjeux de la classification pour les mesures et les représentations de la compréhensibilité de textes. Angelina Aleksandrova et al. Les classifications en linguistique : Problèmes, méthodologie, enjeux, West German University Press, pp.235-251, 2014, 978-3-89966-755-4. halshs-00769135v2

HAL Id: halshs-00769135

https://shs.hal.science/halshs-00769135v2

Submitted on 25 Mar 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Enjeux de la classification pour les mesures et les représentations de la compréhensibilité de textes

Adrien Barbaresi

ICAR UMR 5191, Université de Lyon, ENS Lyon

Résumé Une partie des différents phénomènes linguistiques qu'implique la notion de compréhensibilité peuvent être analysés et mesurés pour classer des textes. Cependant, le classement des textes en fonction d'un indicateur de compréhensibilité, tâche complexe, suppose un découpage conceptuel et méthodologique du savoir linguistique et scientifique, reflété par les notions de lisibilité et de complexité, qui mobilisent différents champs disciplinaires et différents héritages. Ainsi, la première classification pourrait bien avoir pour objet les modèles eux-mêmes, et à travers eux la typologie des critères à retenir. Ensuite, il faut convenir d'une échelle pour observer les phénomènes et de catégories à même de permettre une interprétation fructueuse des résultats. Afin de pouvoir visualiser ces derniers, la question de l'échelle est à nouveau centrale. Toutes les visualisations ne correspondent pas à une stratégie de classification donnée, on peut considérer un texte sous un angle global ou local, unidimensionnel ou multidimensionnel. Enfin, la classification peut avoir un impact sur la réalité, la qualification des phénomènes lors de la production de textes.

Abstract Various linguistic phenomena implied by the notion of comprehensibility can be analysed and measured in order to classify texts. This complex task involves a conceptual and methodological scheme of linguistic and scientific knowledge, as reflected by the ideas of readability and complexity, which deal with different research fields and different frameworks. Thus, the first classification may have to concern the models themselves, and beyond them the typology of criteria that are to be considered. Then, one needs to find a scale to observe the phenomena and a proper categorization to interpret the results. In order to visualize them, the scale issue is once again essential. Not every visualization fits the chosen classification strategy, there are global and local descriptions of the texts, unidimensional and multidimensional ones. Finally, the classification may have an impact on reality, i.e. the identification of phenomena during text production.

Mots-clés: Lisibilité, modélisation, visualisation, interdisciplinarité
Keywords: Readability, scientific modelling, visual analytics, interdisciplinarity

1 Modèles a priori et formules ad hoc

Le domaine de l'assistance à la compréhension et de l'accessibilité des textes regroupe de multiples enjeux, de l'évaluation des textes officiels à

l'établissement de manuels scolaires ou de corpus contrastés pour les apprenants d'une langue étrangère.
Les différents phénomènes linguistiques qu'implique la notion de compréhensibilité peuvent être analysés et mesurés pour classer des textes et/ou pointer des difficultés. Cependant, cette tâche complexe suppose un découpage conceptuel et méthodologique du savoir linguistique et scientifique. La dénomination du faisceau de phénomènes linguistiques observés lors de la détermination de la lisibilité d'un texte recèle en elle-même déjà un choix.
Le classement de textes du plus simple au plus complexe selon des critères linguistiques dans le cadre des études de lisibilité pose deux problèmes majeurs : en amont l'échelle d'observation, les critères à adopter ainsi que leur pondération, et en aval la catégorisation des résultats. Les objectifs de cette opération peuvent varier, de même que les critères et les métriques d'évaluation. On trouve aujourd'hui plusieurs approches concurrentes, qui tirent toutes partie de la numérisation croissante des textes étudiés.
Sadeniemi et al. (2008) par exemple utilisent des tests de compression de fichiers pour évaluer la complexité d'un langage ou d'un texte donné. Or, l'algorithme utilisé (*bzip2*) n'applique pas nécessairement un taux de compression linéaire. Même s'il détecte en principe la redondance d'un texte, il peut être amené à traiter des signaux qui n'ont rien de linguistique et sont liés à l'organisation des données du fichier traité. Les résultats n'ont donc pas de signification immédiate pour un lecteur humain. La simulation s'avère donc être pour le moins imparfaite, si ce n'est simplement inadéquate. Cette approche montre *a contrario* en quoi l'existence de classes raisonnées comme celles de la linguistique est importante pour qualifier un texte en termes de lisibilité.

1.1 Les formules de lisibilité

1.1.1 Enjeux et problèmes

Les formules de lisibilité n'ont rien perdu de leur actualité depuis leur popularisation dans les années 1950. Leur résultat numérique univoque se prête bien à la fixation de seuils permettant une classification par niveaux selon des règles, par la projection du résultat sur un plan défini, un continuum. Elles ont l'avantage indirect de cumuler des informations à travers les phénomènes qu'elles quantifient, souvent de manière linéaire, comme la longueur des phrases ou des mots. Elles peuvent néanmoins

produire des résultats paradoxaux.
Les statistiques textuelles mises en place par le groupe *amazon.com* pour permettre aux usagers d'estimer la lisibilité de certains des livres de leur catalogue peuvent être aisément mise à mal, de même que les délimitations d'âge concernant les livres pour enfants sont sujettes à caution. L'exemple figure 1 illustre à la fois l'intérêt de pouvoir visualiser la difficulté d'un texte et les associations abusives qui peuvent être faites à partir d'indicateurs tronqués ou trop partiels.

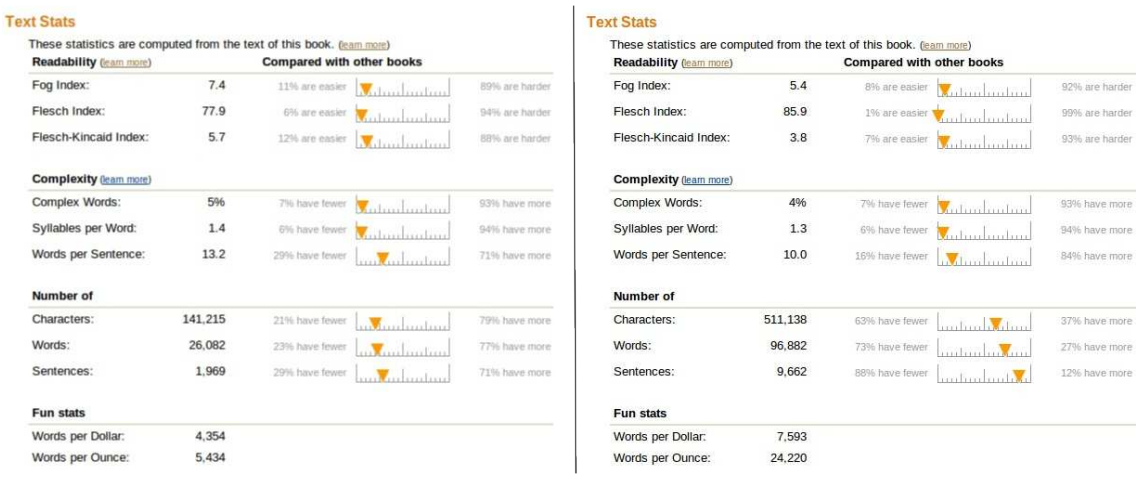

Figure 1: Statistiques textuelles tirées du site amazon.com

Les curseurs censés indiquer la lisibilité (d'après certaines formules) et la complexité du texte (d'après des critères proches des premières) dans les deux premières cases en partant du haut sont positionnés de manière analogue, alors que les deux textes s'avèrent être d'un niveau de difficulté radicalement différent.
Le texte représenté sur la gauche est la traduction anglaise du roman pour enfants *Fifi Brindacier*, de l'auteur suédoise Astrid Lindgren, tandis que le texte à droite est la version originale du roman de William Faulkner *Le bruit et la fureur*, réputé difficile et déconseillé aux enfants en raison de sa richesse et de la violence sociale qu'il dépeint, alors que *Fifi Brindacier* semble être plus facile.

1.1.2 Les différents concepts de lisibilité

L'approche classique en ce domaine, illustrée par les travaux de Dale et Chall (1948), vise à trouver une formule qui, grâce à ses valeurs moyennes issues du travail sur un corpus large, convient à tous les textes. Elle produit un score qui, reporté sur une table, donne par

exemple une accessibilité en terme d'âge ou d'années d'études.
Il faut ajouter à ce concept de « lisibilité classique » les caractéristiques physiques du texte : présentation et taille des caractères, facteurs importants pour les lecteurs humains mais difficilement quantifiables, de même que la densité des idées et des concepts. Avec le développement des sciences cognitives suit dans les années 1980 l'étude de facteurs cognitifs, notamment à l'aune de la notion d'organisation ou de cohérence textuelle (Britton et al. 1982).
Ces analyses détaillées ont un point fort : un classement précis et une assistance détaillée à la rédaction, et un point faible : la généralisation des résultats. Toutefois, il faut reconnaître que les formules linéaires appliquées sans discernement, aussi globales soient-elles, fonctionnent. En effet, les phrases longues comportent généralement une syntaxe complexe, tandis que les mots rares sont souvent des concepts complexes (Kintsch et Miller 1981). Les formules sont donc la plupart du temps en phase avec les propriétés conceptuelles des textes. En revanche, cette efficacité empirique n'aide en rien lors de l'écriture assistée de textes « faciles ».
La visée strictement applicative de ces formules configure également leur typologie des phénomènes et des degrés de lisibilités afférents, qui se trouvent être fortement liés à certains systèmes langagiers et institutionnels. Les formules se situent en rapport d'efficacité avec les catégories qu'elles emploient, leur performance n'apporte aucune connaissance en retour, par exemple pas de typologie textuelle.

1.2 Les modèles linguistiques

On peut donc opposer formules et modèles linguistiques. Les premières proposent une évaluation empirique (éventuellement simpliste) de la lisibilité quand les seconds se fondent sur une approche théorique précise, parfois complexe, qui induit potentiellement des erreurs mais qui permet de réaliser une typologie détaillée des critères.
Dans le premier cas, l'établissement des catégories est tourné vers le domaine applicatif, il se fait *a posteriori* et peut aisément évoluer selon les résultats d'études sur des panels. Dans le second, il se fait plus *a priori*, sur des critères parfois difficiles à mettre en œuvre en situation réelle, comme dans les études en sciences cognitives, où la modélisation s'attache plutôt à illustrer une théorie, par exemple celle du coût de rattachement et du coût en mémoire de Gibson (1998).
À la jonction avec les sciences cognitives, Gibson voit le décodage d'une phrase en termes de ressources nécessaires à sa compréhension. La spé-

cificité de son travail de recherche est de proposer une nouvelle modélisation des relations entre le mécanisme de compréhension du langage et les ressources de calcul disponibles.
Sa théorie s'axe sur l'hypothèse selon laquelle il existe un seuil d'activation au-dessous duquel on ne passe pas à la suite. Il cherche dès lors à déterminer le coût de traitement des groupes ou des propositions avec les outils de la linguistique et des sciences cognitives. Il crée des unités de mémoire censées modéliser l'espace requis par le décodage d'une phrase, en comptant le nombre d'éléments qu'il est nécessaire d'avoir simultanément en tête pour relier des groupes les uns aux autres.
Gibson ne s'intéresse pas qu'à l'espace mais tente également de déterminer le temps nécessaire au décodage d'une phrase à partir d'unités symboliques d'énergie et de mémoire. Il en ressort que les éléments importants, comme les constituants essentiels de la phrase, ne doivent pas être trop distants les uns des autres. S'ils sont au début, la phrase est plus facile. Cette observation rejoint les principes de Hawkins (2004).
Hawkins tente en effet de formaliser le rapport entre complexité et efficacité, partant du principe qu'il y a des raisons au fait qu'un énoncé complexe soit préféré à un autre potentiellement plus simple. Pour lui, la complexité dépend de la structure associée aux mots d'une phrase. Plus de structure signifie plus de singularités linguistiques à traiter.
Sur les bases de la grammaire générative, il développe une théorie raisonnant par unités formelles possédant certaines propriétés qu'il convient d'analyser. La complexité est alors fonction du nombre minimal de propriétés utilisées pour exprimer une proposition donnée. Il formule des maximes (en écho à celles de Grice) qui tentent d'expliquer l'abord de la complexité par les locuteurs. La recherche de la complexité minimale se fait d'après lui selon le principe « exprimer le plus possible avec le moins possible» et «l'exprimer le plus tôt possible » . Il s'agit de minimiser les formes employées et de maximiser les possibilités de rattachement et de rapprochement des éléments les uns aux autres (« *maximize on-line processing* »).
Partant, sa démarche se trouve à mi-chemin entre une approche logique et une autre cognitiviste, elle s'appuie sur les dernières théories de Chomsky, tandis que celle de Gibson constitue une approche fine qui aborde une complexité dynamique, qui s'établit au fur et à mesure que la phrase est comprise. Elle va dans le sens des études sur la progression textuelle.

1.3 Approches hybrides et apprentissage artificiel

Enfin, il existe des approches hybrides qui, partant d'un modèle linguistique, laissent le soin à des classifieurs d'établir une typologie voire une hiérarchie des critères à retenir (selon un fonctionnement analogue à la classification automatique des e-mails par exemple, qui permet de discerner les messages importants et le spam). Il est même possible d'obtenir une évaluation quantitative de la distribution des catégories et donc de faire évoluer le classement à la fois du point de vue des critères et du point de vue des résultats.
Le modèle de langue retenu peut découler d'une analyse experte, proprement linguistique, ou être établi automatiquement à partir d'un corpus d'entraînement, c'est-à-dire à partir de grandes quantités de textes supposés analogues au corpus à traiter. Quand le modèle établi n'est pas capable d'inférer une règle compatible avec la partie du texte à analyser, on parle de perplexité, désignant alors une forme de complexité liée à l'imprévu et aux structures rares.
L'entraînement sert également à étalonner la classification des critères, qui se fait selon l'adéquation du résultat à des catégories prédéfinies. Le dépistage des traits saillants passe par des coefficients de corrélation qui ont pour but de dévoiler des régularités statistiques à grande échelle. L'impact quantitatif d'un phénomène linguistique sur l'opération de classification peut ainsi être identifié, alors que ce n'est pas systématiquement le cas lors d'une analyse « experte ».
Par exemple, Kate et al. (2010) déterminent à l'aide du logiciel Weka l'algorithme de régression le plus efficace. Ils comparent en outre un panel « expert » car spécialisé dans les métiers de la langue et un panel « naïf » car étranger aux concepts linguistiques, et trouvent que certains algorithmes de tri surpassent le panel « naïf » en termes d'efficacité. Cela étant, ils prennent comme référent les décisions prises par les « experts », alors même que l'on peut estimer qu'il ne faut aucune expertise autre que celle de savoir lire pour estimer la lisibilité d'un texte.

1.4 Bilan

Ces trois exemples montrent qu'il existe des modèles en profondeur de la langue et des approches applicatives dont le versant opératoire évacue les concepts linguistiques établis. Il y a donc plusieurs cadres disciplinaires différents en jeu, et deux principaux points de vue sur la langue et son examen : d'un côté une approche au fil du texte reflétant une ap-

proche locale des phénomènes et de l'autre un bilan souvent chiffré qui donne une image globale de la lisibilité. La question de l'échelle est la résultante directe d'un positionnement disciplinaire, elle décide d'une typologie de phénomènes qui seront retenus pour l'analyse.

2 Échelles subjectives et grilles de critères interopérables

2.1 Linéarité des catégories

Les catégories de difficulté ne sont pas nécessairement bien définies ou homogènes, ce qui rend leur classification sujette à caution, comme l'expliquent Heilman et al. (2008), qui distinguent notamment trois échelles de mesure : nominale, ordinale et par intervalle. En l'adaptant à la lisibilité, on peut dire que « simple », « moyennement difficile » et « difficile » par exemple sont des catégories appartenant à échelle nominale. Il n'y a pas de rapport immédiatement quantifiable entre elles.
L'adoption des degrés d'un système scolaire relève d'une échelle ordinale: si l'ordre choisi paraît quantifiable, les valeurs sur l'axe d'évaluation ne se trouvent pas nécessairement à intervalles réguliers. C'est une propriété importante de ce type de classification, qu'il convient dès lors d'interpréter avec une certaine marge d'erreur. Enfin, l'établissement d'un indice dédié explicitement à l'estimation de la lisibilité représente une échelle où les intervalles sont pertinents.
Ainsi, l'établissement de catégories d'une nature ou d'une autre peut entraîner différentes conséquences méthodologiques, tant lors du classement que lors de l'évaluation.

2.2 Définition des catégories et stratégie de classification

L'apprentissage en séquences à l'école et l'établissement de référentiels de compétences facilite l'étalonnage des mesures de même que leur réalisation, en particulier dans le cas de l'apprentissage d'une langue étrangère. Cependant les normes valant pour un certain système scolaire ne sont pas transposables telles quelles dans un autre.
Le découpage en niveaux de lisibilité se fait d'après une échelle souvent liée à une tradition scientifique, à des systèmes institutionnels d'évaluation et/ou à l'expérience amassée dans un domaine particulier. L'évaluation des textes utilisés lors de l'élaboration des manuels scolaires par exemple recoupe ces trois préoccupations.

Ott (2009) note ainsi l'existence de différentes échelles. Aux États-Unis, où les niveaux de 1 à 17 suivent les années d'enseignement du primaire au supérieur, au Royaume-Uni où l'on adopte une échelle à cinq niveaux. Autre exemple, le Cadre européen commun de référence pour les langues (CECR), plus récent, distingue trois niveaux comportant à chaque fois deux sous-niveaux.
Cette approche n'est pas partagée par tous les chercheurs, vor der Brück et al. (2008) évaluent un outil d'aide à la rédaction en demandant à un panel de classer des textes en allemand sur une échelle axée sur les impressions subjectives des lecteurs, là où le CECR propose une grille de compétences relatives à des situations précises de communication, comme l'échange d'informations dans un cadre routinier pour le niveau élémentaire (A2). Ce canevas, certes lâche, a l'avantage de couvrir de nombreuses situations et de pouvoir être généralisé à plusieurs langues.
Étant donnés la large diffusion du CECR et son adoption par la plupart des organismes de certification, il semble être un choix porteur. De plus, comme l'explique François (2009), il permet de réfléchir aux schémas employés lors de l'élaboration de séquences de cours ainsi qu'au calibrage des nombreux manuels qui suivent ce format. L'utilisation du passé simple en français devient alors l'élément d'une stratégie de classification sur une échelle de lisibilité.

2.3 Une grille interopérable

L'analyse de la lisibilité des textes peut et doit de préférence s'inscrire dans une opération d'enrichissement, qui ajoute des strates de connaissances à un terrain d'expérience, un corpus, afin de faciliter sa réutilisation et son insertion dans une économie de la preuve scientifique.
Daoust et al. (1996) remarquent ainsi qu'au-delà d'un indice numérique, on peut effectuer une classification fondée sur des « caractéristiques spécifiques » :

« Il est vite apparu que la production d'un indice n'était pas le plus important. Ce que permet l'application et ses extensions possibles, c'est de repérer des textes possédant des caractéristiques spécifiques pour élaborer des instruments d'enseignement ou d'évaluation. » Daoust et al. (1996), p. 231.

Réduire la difficulté, la lisibilité ou la complexité à un chiffre masque la nature diverse de variables linguistiques qui opèrent conjointement et

qu'un marquage au fil du texte retraçant la linéarité de la lecture permet de faire ressortir.
Il est donc souhaitable d'aboutir à un étiquetage « partiel et partial » voire même « inéquitable » de surcroît (Habert et al. 1997), au sens où l'on s'intéresse aux phénomènes liés à la compréhensibilité, en laissant les autres « dans l'ombre », ce qui est encore une stratégie de classification pour ne pas alourdir inutilement le marquage qui permet de reconstituer l'analyse.

3 Représentation et évaluation des résultats

Les choix qui sous-tendent la stratégie de repérage prédéfinissent également les critères de son évaluation, dans un certain cadre disciplinaire (cf introduction). Outre l'évaluation des indicateurs de lisibilité, on peut s'interroger sur l'image que nous devons avoir de ces phénomènes, ainsi que sur la représentation que l'on peut en donner.

3.1 Perspectives globale et locale

L'échelle de la représentation peut différer de celle de l'observation, on peut effacer les différences locales sous un résultat global ou les conserver. La perception de la difficulté ne se fait pas nécessairement à la même échelle non plus : elle peut être locale, à un endroit donné du texte, ou globale et systématique. La classification pourrait donc se faire non sur un axe, mais en plusieurs dimensions, au fil du texte, et selon des profils variables, ce qui permettrait de profiter du marquage des difficultés pour rendre visible la progression du lecteur dans le texte.
L'évaluation ne se ferait donc pas directement à partir d'une classification de textes par rapport à un standard (un axe), bien que celle-ci reste possible à partir de valeurs numériques calculées à partir du marquage, mais plutôt à partir de la justification de la mise en relief des passages difficiles.

3.2 Unidimensionnalité et multidimensionnalité

La multiplicité des variables conduit à une multidimensionnalité des résultats, et ce qu'affirme Legay sur l'articulation entre les variables dans la mesure s'avère pertinent dans ce cas, avec des variables linguistiques.
Biber (1992) est partisan d'une analyse multi-dimensionnelle, car il estime les phénomènes à observer ne sont pas liés sur un même niveau lin-

guistique.

« Given this multidimensional nature of discourse complexity, a complete description of a text (or register) cannot focus exclusively on the degree of complexity; rather, the complexity of a text is described here within a five-dimensional space, so that texts are complex in different ways as well as to differing extents. » Biber (1992), p. 149.

Dans ce cas l'évaluation peut passer par des tests de corrélation entre un ensemble de phénomènes et leurs indicateurs qui réduisent (au mieux de manière transparente) les multiples dimensions de l'analyse. Deux problèmes subsistent alors. D'une part, la pondération et l'articulation des critères, l'équilibre à trouver, auquel la notion de profil permet d'apporter plus de profondeur : on ne lit pas forcément de la même manière à tous les âges par exemple. D'autre part, la représentation de l'analyse : comment donner à voir un grand nombre de phénomènes opérant de manière simultanée et parfois concurrente ?

3.3 L'apport d'une visualisation construite

L'œil humain classifie ce qu'il voit, souvent intuitivement. Il peut dès lors être productif de construire une visualisation de ce que l'on souhaite observer, afin tout d'abord de contrôler le fonctionnement de ses instruments, c'est-à-dire par exemple un repérage conforme des phénomènes, mais également de parvenir à des découvertes qui seraient restées dissimulées jusque là sous la pro- et la confusion de détails.
La visualisation est un domaine scientifique à part entière, où la catégorisation des phénomènes à représenter joue un rôle primordial. Quelques notions élémentaires de visualisation sont ainsi énumérées par Bertin (1967) : pour éviter les recoupements factices, il faut par exemple adopter un code couleur approprié et jouer sur les nuances plus que sur des différences de teintes, qui supposent une hétérogénéité.
À travers la visualisation un ré-échelonnement s'opère. Cela permet de prendre du recul et ainsi de détecter des problèmes de méthode à plus large échelle, de regrouper intuitivement des phénomènes jusque là disjoints, ou d'apporter une autre lecture des résultats.
Il n'en demeure pas moins que cette démarche repose sur une construction, en l'espèce une reconstruction du texte qui s'effectue là encore selon différentes traditions scientifiques. À une approche issue du traitement automatique des langues, centrée sur des techniques de repérage et des

métriques d'évaluation, s'oppose une approche utilisée par des chercheurs en informatique et en visualisation, qui est une adaptation de techniques de visualisation courantes à cet objet de recherche, comme c'est le cas pour Oelke et al. (2010).

4 Conclusion

4.1 Le découpage disciplinaire et ses enjeux

Selon que l'on parle de complexité, et éventuellement de complexité linguistique, lexicale ou syntaxique par exemple, de lisibilité ou de compréhensibilité, on mobilise différents concepts liés à des disciplines comme les sciences cognitives et les sciences du langage ainsi que leurs sous-ensembles, comme la psycholinguistique ou la linguistique textuelle. Cette distinction terminologique recoupe également différents champs de connaissance et différentes méthodes qui sous-tendent l'espace disciplinaire choisi. Dès lors, choisir un de ces termes implique d'opter pour un certain découpage conceptuel et méthodologique résultant d'un héritage scientifique.

4.2 De la mesure à la classification

Le contexte disciplinaire du traitement automatique des langues (TAL) ne se réduit pas à un passage sur les textes, un simple traitement, même si certaines études s'inscrivent dans cette perspective. Il s'agit également de tirer des conclusions sur le langage :

« Le TAL n'est pas qu'un outil de traitement (c'est-à-dire de transformation) mais il est aussi un moyen d'observation du matériau langagier. Il s'agit à la fois d'un microscope qui permet d'observer de petits phénomènes, et d'un macroscope car il permet d'observer au sein d'une grande masse de données. » Lebarbé (2010, p. 60.

L'image du microscope et du macroscope peut être étendue à une certaine définition de la mesure, qui n'a pas pour but simplement de donner une information mais aussi d'étayer un raisonnement et de soutenir une logique de découverte scientifique.
En ce sens, la mesure et la classification ne sont pas nécessairement des pôles fonctionnant indépendamment l'un de l'autre, dans certains cas elles peuvent être étroitement liées, comme dans la classification auto-

matique, qui calcule en fait une distance entre les textes. Plus globalement, c'est le constat que fait Dagognet (1994), selon lequel la mesure indique des niveaux, hiérarchise, jette sur une grille quantitative, tout en ne se séparant pas d'un ensemble de variables. Parce qu'il projette, l'instrument de mesure est instrument de classification, affirme-t-il. Et de fait, l'analyse de la lisibilité, en retournant un chiffre, sert souvent à classer des textes, malgré les éventuels problèmes de linéarité des axes évoqués plus haut.
Aux mesures de lisibilité s'oppose un autre type de mesure, plus construit et moins directement réutilisable, une mesure qui s'éloigne de la simple fonction de raccourci vers le réel pour devenir elle-même une intégration complexe d'indicateurs, dont la pertinence passe par un classement et un tri de ces derniers. C'est cette opération que décrit Jean-Marie Legay quand il affirme :

« La mesure s'est peu à peu éloignée de son objet, s'est écartée d'un objectif direct et immédiat; elle n'est plus suivie de conclusions évidentes; elle entre elle-même dans un système complexe de connaissances. » Legay (1994), p. 146.

Legay constate un déplacement du problème central de l'opération de mesure, il ne s'agit plus de réduire un objet à un signe, mais de choisir un certain nombre de variables et de les articuler entre elles. Peut-être est-ce là une acception possible de la notion d'indicateur : une intégration raisonnée de données expérimentales.

4.3 Une typologie doublée d'une triple opération de classification

La sélection des variables linguistiques qui doivent rendre compte dans leur ensemble de la compréhensibilité de textes s'axe sur une typologie de phénomènes jugés pertinents ou non, éventuellement au regard de différentes catégories linguistiques comme la morphologie, la syntaxe ou la sémantique. Cette typologie est fonction de et appuie à la fois trois opérations de classification à trois niveaux différents.
La première opération est de nature méta-linguistique, elle a pour objet les modèles eux-mêmes, c'est-à-dire le cadre théorique, la méthode employée, les objectifs qui prévalent et au regard desquels la démarche scientifique sera évaluée. Ce positionnement au sein d'une classification du savoir linguistique implique une certaine idée de la typologie des critères à retenir, qui permet l'étude de traits, de données empiriques. Il dé-

cide également de l'usage privilégié d'une terminologie (cf *supra*). Cependant, ce choix peut être implicite, sans que l'on soit forcément au clair sur les modèles sous-jacents, et donc receler une part de syncrétisme.
La seconde classification prend la forme d'un tri qui découle de la première, elle définit un terrain d'application concernant le repérage des traits linguistiques. Il faut en effet convenir d'une échelle d'observation (par exemple la phrase, le paragraphe ou le texte) et d'une catégorisation des phénomènes à même de permettre une interprétation fructueuse des résultats.
Enfin, la dernière a trait à la représentation des résultats. Elle dépend étroitement de la stratégie de repérage choisie. On peut ainsi considérer un texte sous un angle global ou local, unidimensionnel ou multidimensionnel. On peut réduire les variables observées pour les projeter sur un axe, comme dans le cas des scores de lisibilité, ou désirer obtenir une représentation soucieuse de leur diversité. Étant la plus manifeste, cette opération est la plus « publique », elle peut avoir un impact sur la réalité, la qualification des phénomènes et le changement des pratiques lors de la production de textes.

4.4 Impact sur la réalité observée

Enfin, la recherche dans ce domaine doit s'accompagner d'une réflexion sur l'impact des résultats, en gardant en point de mire leur valeur normative. Une transformation de la réalité par la mesure est possible, selon ce que Dagognet (1994) dit de l'histoire de la physique : ce qui a été mesuré (tel corps fond à tant de degrés) sert à son tour d'indicateur, « la mesure va jusqu'à transformer le mesuré en mesurant ! »
Ce cas de figure est d'autant plus prégnant dans le cas de la classification selon une norme : l'évolution des pratiques collectives afin de rédiger des textes plus simples en est un exemple. Un système choisi donne aux textes écrits une certaine tournure, que l'on peut éventuellement détecter.

4.5 « Croiser les approches »

Face à d'éventuelles réécritures qui pourraient tromper le dispositif de mesure sans améliorer la lisibilité des textes, le meilleur rempart reste la finesse de l'indicateur lui-même. Pour ce faire, on peut imaginer de « croiser les approches » par une approche soucieuse des avancées dans les domaines concernés et une démarche d'intégration des indicateurs de

compréhensibilité quelque soit leur nature, afin de « munir les données attestées d'annotations fines, multiples, permettant de progresser vers les régularités sous-jacentes » (Habert et Zweigenbaum 2002).
Il s'agit de tirer parti de la richesse des différents modes opératoires décrits plus haut, que ce soit dans l'établissement d'un référentiel (la validation par un panel ou des textes déjà labellisés) ou dans l'évaluation des résultats (coefficients de corrélation, pertinence du classement), avec pour objectif d'enrichir les textes (Barbaresi 2011).
Alors que le résumé chiffré d'un texte empêche tout « effet de zoom » et toute déconstruction, procédés qui permettraient de cerner des passages réputés difficiles par exemple, un marquage des difficultés au fil du texte permet de reconstruire différents indicateurs en aval. L'évaluation de cette procédure ne se ferait pas par la classification globale d'un texte, mais plutôt en demandant à un panel de surligner les passages difficiles. Cette démarche présente l'intérêt de s'intéresser à la manière dont les « lecteurs naïfs » lisent et conçoivent des stratégies de repérage qui ne sont pas forcément prévisibles.
L'idéal étant de garder la trace des différents procédés afin de pouvoir recombiner les annotations pour en tirer de nouvelles connaissances, dans un instrumentation qui crée de nouveaux observables (Valette 2008).
De la même manière, on peut imaginer de croiser différentes traditions de visualisation pour proposer une interface dynamique et réactive qui propose différentes échelles d'observation en prêtant attention aux besoins de l'utilisateur.

Je remercie Benoît Habert pour ses conseils et sa relecture attentive.

Références

Barbaresi, Adrien (2011): *Approximation de la complexité perçue, méthode d'analyse.* Dans : Actes TALN'2011/RECITAL, vol. 2, 229–234.
Bertin, Jacques (1967): *Sémiologie graphique. Les diagrammes, les réseaux, les cartes.* Paris, La Haye : Mouton, Gauthier-Villars.
Biber, Douglas (1992): *On the complexity of discourse complexity : A multidimensional analysis.* Discourse Processes, 15(2), 133–163.
Britton, Bruce K. / Glynn, Shawn M. / Meyer, Bonnie J. / Penland, M. J. (1982): *Effects of text structure on use of cognitive capacity during reading.* Journal of Educational Psychology, 74(1), 51–61.
vor der Brück, Tim / Helbig, Hermann / Leveling, Johannes (2008): *The Readability Checker DeLite (Rapport technique)*. Hagen : FernUniversität in Hagen, Intelligente In-

formations- und Kommunikationssysteme.
Dagognet, François (1994): *Pourquoi nous persistons a valoriser la mesure ?* Dans: Beaune [Ed.] (1994): *La mesure, instruments et philosophies.* Seyssel : Champ Vallon, 81–85.
Daoust, François / Laroche, Léo / Ouellet, Louise (1996): *Sato-Calibrage : Présentation d'un outil d'assistance au choix et a la rédaction de textes pour l'enseignement.* Revue québécoise de linguistique, 25(1), 205–234.
François, Thomas (2009): *Modèles statistiques pour l'estimation automatique de la difficulté de textes de FLE.* In Actes TALN/Récital. Senlis, ATALA.
Gibson, Edward (1998): *Linguistic complexity : Locality of syntactic dependencies.* Cognition, 68(1), 1–76.
Habert, Benoît / Nazarenko, Adeline / Salem, André (1997): *Les linguistiques de corpus.* Paris : Armand Colin.
Habert, Benoît / Zweigenbaum, Pierre (2002). *Régler les règles.* TAL, 43(3), 83–105.
Hawkins, John (2004): *Efficiency and complexity in grammars.* New York : Oxford University Press.
Heilman, Michael / Collins-Thompson, Kevyn / Eskenazi, Maxine (2008): *An Analysis of Statistical Models and Features for Reading Difficulty Prediction*. In *Proceedings of the third workshop on innovative use of nlp for building educational applications,* Stroudsburg, PA, USA : Association for Computational Linguistics, 71–79.
Kate, Rohit J. / Luo, Xiaoqiang / Patwardhan, Siddharth / Franz, Martin / Florian, Radu / Mooney, Raymond J. / Roukos, Salim / Welty, Chris (2010). *Learning to Predict Readability using Diverse Linguistic Features*. Dans : *Proceedings of the 23rd international conference on computational linguistics.* Stroudsburg, PA, USA : Association for Computational Linguistics, 546–554.
Kintsch, Walter / Miller, Jim R. (1981). *Readability : A view from cognitive psychology*. Teaching Research Reviews, 220–232.
Lebarbé, Thomas (2010). *Fonctions interdisciplinaires intrinsèques et extrinsèques du traitement automatique des langues – synthèse d'habilitation à diriger des recherches*. Université Grenoble 3
Legay, Jean-Marie (1994). La mesure dans les systèmes complexes. Dans: Beaune [Ed.] (1994): *La mesure, instruments et philosophies.* Seyssel : Champ Vallon, 81–85.
Oelke, Daniela / Spretke, David / Stoffel, Andreas / & Keim, Daniel A. (2010). *Visual readability analysis : How to make your writings easier to read.* Dans : *VAST 10 : IEEE Conference on Visual Analytics Science and Technology*. Salt Lake City : IEEE, 123–130.
Ott, Niels (2009). *Information Retrieval for Language Learning : An Exploration of Text Difficulty Measures*. Mémoire de Master non publié. Université de Tübingen, Seminar für Sprachwissenschaft.
Sadeniemi, Markus / Kettunen, Kimmo / Lindh-Knuutila, Tiina / Honkela, Timo (2008). *Complexity of European Union Languages : A comparative approach*. Journal of Quantitative Linguistics, 15(2), 185–211.
Valette, Matthieu (2008). *Pour une science des textes instrumentée*. Syntaxe et sémantique, 9, 9–14.