

DOSSIER RECHERCHES ACTIONS (4)

GENÈSE DU TEXTE : SAVOIRS

Claire DOQUET

Cet article fait suite à celui du numéro précédent (AL n°47, sept.94, p.99) qui présentait les variables que nous avons retenues pour analyser les processus d'écriture des textes. Cette analyse correspond au deuxième volet de la recherche Genèse du texte de l'INRP, où le logiciel est "utilisé comme auxiliaire de recherche sur la naissance du texte, lieu où se forge et se dévoile la constitution progressive de la compétence scripturale" (AL n°37, mars 92, p.76). Nous présentons ici les variables retenues pour l'analyse des textes finaux produits par les élèves.

L'objectif de cette évaluation de la production est de dépasser les critères formels d'orthographe et de correction syntaxique pour observer le fonctionnement de la langue écrite. Cette observation se fait de façon automatique grâce au logiciel **Analyse de textes** qui recense un certain nombre de variables linguistiques. Les avantages de l'utilisation du traitement automatique sont évidents : la rapidité d'analyse du logiciel permet la prise en compte d'un grand nombre de textes, ce qui ne serait pas possible si chacun d'entre eux devait être lu et annoté à la main. Ses inconvénients sont liés aux limites de l'ordinateur qui est évidemment incapable de tirer d'un texte son sens et d'apprécier le jeu avec la langue que les lecteurs y recherchent. Le premier travail à effectuer est donc le passage de critères subjectifs (ce texte est bien écrit, il est travaillé, etc.) à des critères objectifs et mesurables par le logiciel (la proportion d'adjectifs dans ce texte est significative d'une attention particulière de l'auteur).

Cet exemple peut sembler réducteur mais c'est bien ce type de critères que mesure le logiciel et qu'il faut ensuite évaluer. Pour éviter les interprétations erronées, et puisqu'il est évident que la qualité d'un texte ne dépend pas d'un seul de ces critères, nous les avons regroupés en catégories plus larges prenant en compte plusieurs critères mesurables. Ces catégories sont les suivantes :

- les thèmes employés,
- la variété de construction des phrases,
- l'organisation des phrases,
- l'organisation du texte
- les marques de l'attention portée par le scripteur au matériau qu'il utilise : la langue écrite.

LES THÈMES

Il s'agit moins ici de s'intéresser à ce qu'évoquent les textes qu'à la manière dont ils l'évoquent et à la répartition des différents thèmes dans le texte. Nous avons cherché à répondre à deux questions :

- quels sont les instruments de l'évocation des thèmes et comment ceux-ci se répondent-ils dans le texte ?
- par le filtre de la disposition des thèmes, en quoi chaque paragraphe constitue-t-il une reprise du texte lui-même ?

Le recouvrement des thèmes

La première préoccupation concerne le vocabulaire illustrant les thèmes dominants du texte. Il s'agit de savoir si, grâce à l'emploi de mots polysémiques, on peut considérer qu'un thème est évoqué sans être explicitement dénoté. En d'autres termes : certains mots sont-ils porteurs de plusieurs thèmes et ont-ils de ce fait une fonction de reprise thématique, même si cette reprise passe inaperçue aux yeux

du lecteur ? Cette notion de reprise est un instrument de la cohérence du texte. Elle engendre aussi des commentaires riches : n'est-ce pas elle que recherchent les élèves chargés d'une explication de texte ?

Le recouvrement thématique nous intéresse à double titre :

- Comme cela a déjà évoqué, le rappel d'un thème déjà apparu par un mot dont le sens premier se rattache à un autre thème participe de la cohérence du texte. À travers un vocabulaire polysémique, c'est la mise en relation implicite de deux thématiques qui se fait jour. Si deux thèmes distincts sont évoqués séparément, le fait que certains mots soient communs à l'un et l'autre induit des rapprochements et est sujet à réflexion pour le lecteur.

- Du point de vue du scripteur, on peut penser que les connotations sémiques portées par un mot appartenant au thème principal vont être déterminantes dans le choix de thèmes secondaires. Dans ses ateliers d'écriture, Claudette Oriol-Boyer a beaucoup travaillé à partir de la polysémie des mots : demandant de dresser la liste des mots se rapprochant par leur sens de la mer, elle obtient, entre autres sable, vagues ou écume, l'adjectif jaune (donné comme couleur du sable) et le nom grève. Tout le jeu de l'écriture consiste ensuite à exploiter la polysémie : *dans une commune maritime, la grève s'enlise à cause des jaunes...*

Qui n'a pas, en relisant un texte, aperçu entre les lignes la possibilité d'une autre interprétation que celle explicitement voulue par l'auteur ? De ce fait, comment affirmer que les thèmes qui se suivent sont là par une volonté préalable de l'auteur et que la polysémie des premiers mots n'a pas participé à leur émergence ? Pour travailler ces questions, nous demandons au logiciel, après détermination des trois thèmes dominants de chaque texte, de recenser quels mots renvoient à plusieurs de ces trois thèmes. À ces mots est affecté un coefficient et l'on obtient après analyse le taux de "recouvrement" des thèmes par des mots identiques.

La structure thématique du texte

Le deuxième aspect de l'étude des thèmes concerne leur hiérarchie dans le texte et dans chacun des paragraphes. Il s'agit de savoir si les paragraphes conservent la hiérarchie globale des thèmes ou si celle-ci évolue en cours de texte. Dans le premier cas, le texte sera très structuré, l'auteur prenant soin d'évoquer dans chaque paragraphe les thèmes dominants dans un ordre décroissant correspondant à celui du texte. C'est le schéma typique de la dissertation : l'introduction pose les thèmes, le développement les traite, la conclusion les élargit ; à l'intérieur de chaque partie du développement, ces thèmes sont évoqués à nouveau. Au contraire, le deuxième cas représente des textes dont la structure est moins classique, le paragraphe ne se posant plus comme une unité représentative de la globalité du texte.

LA VARIÉTÉ DE CONSTRUCTION DES PHRASES

La variété de construction des phrases est un témoin de l'aisance du scripteur à manier les structures syntaxiques de la langue. Nous avons déterminé quatre types de phrases :

- Les phrases indépendantes et unipropositionnelles : *Le petit chat est mort* (Molière, **L'École des Femmes**).

- Les phrases comportant plusieurs propositions indépendantes liées par des conjonctions de coordination ou de la ponctuation : *Je suis en effet précieux pour vous, car je tuerai Alexandre* (Musset, **Lorenzaccio**).

- Les phrases comportant plusieurs propositions liées par des conjonctions de subordination uniquement : *La vue de cette femme qui l'avait tant aimé fit trembler le bras de Julien d'une telle façon, qu'il ne put d'abord exécuter son dessein* (Stendhal, **Le Rouge et le noir**).

- Les phrases comportant plusieurs propositions liées par des conjonctions de subordination et de coordination : *Il regardait Brulette qui s'était appuyée contre une chaise et qui avait la figure tournée du côté du mur* (Georges Sand, **Les Maîtres Sonneurs**).

La proportion de chacun de ces quatre types de phrases devrait nous renseigner sur la complexité du texte ainsi que sur le mode d'énonciation. Dans la littérature des siècles passés, le théâtre, à l'instar des exemples donnés chez Molière et Musset, fourmille de phrases des deux premiers types alors qu'elles sont relativement difficiles à trouver dans le roman, illustré par les exemples de Sand et Stendhal. Il serait dangereux d'inférer ces constatations à la production écrite actuelle, qu'il s'agisse de littérature ou de tout autre chose : on sait pas exemple que Marguerite Duras emploie volontiers dans ses romans des phrases simples où les propositions sont reliées par des conjonctions de coordination. Toutefois, le modèle de l'écrit dans un cadre scolaire reste largement empreint de la production du XIX^{ème} siècle et il est légitime de se demander si les structures phrastiques employées correspondent à une volonté de l'auteur pour faire écrit, même si cette notion est aujourd'hui largement remise en cause dans la production littéraire.

Pour affiner cette typologie de construction de phrases, nous recensons également un certain nombre de caractéristiques statistiquement rares et qui semblent relever d'un effort d'expression particulier :

- Parmi les phrases complexes, celles qui commencent par une subordonnée : *Que tu viennes du ciel ou de l'enfer, qu'importe !*
- Parmi les pronoms relatifs employés, la proportion de pronoms relatifs autres que qui et que, qui sont les plus employés.
- Parmi les conjonctions de subordination, la proportion de conjonctions peu employées d'ordinaire.

L'ORGANISATION DES PHRASES

Par organisation des phrases, nous entendons la manière dont celles-ci sont structurées et le type de formulation employé. Il s'agit de qualifier la structure des groupes nominaux qui peuvent être des groupes simples (le petit chat) ou des groupes complexes (le petit chat de ma grand-mère qui habite Lyon). Cette distinction est mise en évidence par Bronckart (1985) sous le terme de densité syntagmatique : plus celle-ci est forte, plus le nombre de qualifiants (petit, de ma grand-mère qui habite Lyon) est important par rapport aux noms noyaux (chat) directement rattachés au verbe.

Nous calculons :

- La densité syntagmatique, soit le rapport du nombre de qualifiants et le nombre de noms noyaux.
- La densité verbale, soit le rapport entre le nombre de verbes et le nombre de mots.

L'ORGANISATION DU TEXTE

Type de texte et intention de l'auteur

Nous avons mis au point une typologie des textes à partir de la typologie des discours de J.P. Bronckart (*Le Fonctionnement des discours*, Delachaux et Niestlé, 1985). Elle se base sur les intentions de l'auteur sur son lecteur. Tout texte est caractérisable selon quatre axes et peut ainsi être qualifié de :

- impliqué ou autonome
- conjoint ou disjoint
- intégré ou délégué
- impositif, expositif ou propositif.

Cette typologie, ainsi que les 27 variables que Bronckart utilise pour la déterminer, ont déjà été exposées dans le n°38 des AL par Thierry Opillard (pp. 56-59) et Robert Caron (pp. 66-73), nous ne les réexpliquerons pas ici.

Dans le questionnaire à l'élève (Cf. AL n°47, p.90, "*Caractérisation des conditions de production*") figurent ses intentions sur son lecteur : il répond indirectement à la question du type de texte qu'il a voulu produire. Cette réponse est comparée au type de texte réellement produit, déterminé automatiquement par le logiciel, et nous savons donc si les intentions de l'auteur sont ou non en adéquation avec le résultat.

Les temps verbaux

Les temps employés structurent l'ensemble du texte et marquent la prise en compte de la chronologie. En outre, les temps des verbes employés nous intéressent en tant que leur diversité est témoin de l'aisance du scripteur à manier les caractéristiques temporelles et aspectuelles des temps et modes. L'indicatif étant le mode le plus fréquemment employé, nous observons quel temps de l'indicatif domine dans le texte et dans quelle proportion sont représentés les autres modes. Tous modes confondus, le nombre de temps verbaux donnera une indication de la diversité de l'emploi des temps et modes.

Mode d'organisation du propos

Indépendamment des champs sémantiques qu'il comporte, le texte peut s'organiser autour de :

- la transformation,
- la description,
- le raisonnement,
- la chronologie.

Il s'agit de déterminer quels modes dominent et comment les modes se répartissent au cours du texte. Pour chaque texte, nous relevons les deux modes d'expression dominants et les comparons aux modes dominants de chaque paragraphe. Il devient donc possible d'observer dans quelle proportion les paragraphes reprennent ces modes dominants et également la diversité de la disposition des modes dans les différents paragraphes. La comparaison du mode dominant du premier et du dernier paragraphes nous renseignera aussi sur le texte lui-même : clos si le même mode domine, évolutif si l'on passe d'un mode à l'autre.

Argumentation, type de verbe

Pour compléter ces informations, nous observons aussi dans quelle proportion sont représentés les organisateurs de l'argumentation, du temps et de l'espace, ainsi que le type d'organisateur dominant.

Le linguiste John Austin (Quand dire, c'est faire, Seuil, 1991) a classé les verbes en différentes catégories basées sur l'action que produit l'énonciation elle-même. L'exemple type est le verbe verdictif : quand le Maire dit, dans la situation adaptée, "je vous marie", le simple fait de dire constitue l'acte de mariage du couple concerné. En référence aux catégories verbales élaborées par Austin, nous observons quel type de verbe domine et dans quelle proportion sont représentés les 5 types de verbes.

LE SOUCI DE L'ÉCRITURE

Dans cette partie, il s'agit d'évaluer le degré d'attention portée par le scripteur aux mots et formes syntaxiques employés. Il sera particulièrement intéressant de comparer ces caractéristiques à celles des processus d'écriture puisque l'on peut penser qu'un scripteur effectuant dans son texte beaucoup de modifications produit un texte plus abouti, y compris formellement, qu'un scripteur de premier jet.

Pour tenter d'observer de façon automatique, dans le texte final, les marques d'une attention particulière du scripteur pour l'écriture, nous avons déterminé des caractéristiques statistiquement rares et visant à produire un effet particulier.

Agencements syntagmatiques remarquables

Sachant que l'ordre le plus fréquent des mots dans la phrase est la disposition sujet - verbe - complément, nous nous intéressons à la proportion de phrases commençant par :

- Un adverbe : "Ainsi se débattait sous l'angoisse cette malheureuse âme." (Hugo, *Les Misérables*).
- Une préposition : "Par l'art seulement nous pouvons sortir de nous, savoir ce que voit un autre de cet univers qui n'est pas le même que le nôtre, et dont les paysages nous seraient restés aussi inconnus que ceux qu'il peut y avoir dans la lune." (Proust, *Le Temps retrouvé*).

- Une conjonction de subordination : "*Si son enfance se fût écoulée dans l'arrière-boutique d'un quartier marchand, elle se serait peut-être ouverte alors aux envahissements lyriques de la nature, qui, d'ordinaire, ne nous arrivent que par la traduction des écrivains.*" (Flaubert, *Madame Bovary*).
- Un adjectif : "*Mystérieuse Moi, pourtant, tu vis encore !*" (Valéry, *La Jeune Parque*).

Les formes syntaxiques particulières telles que les sujets inversés et les formes passives seront prises en compte, relativement au nombre de verbes que contient le texte.

Ponctuation

Nous recherchons les signes de ponctuation semblant marquer une intention particulière. Le signe de ponctuation minimal et obligatoire dès qu'un texte comporte plus d'une phrase est le point : il n'est pas pris en compte. La virgule, le point d'exclamation et le point d'interrogation doivent être pris en compte mais dépendent surtout de la forme de la phrase, relevée par ailleurs. En revanche, les deux points et points de suspension, et plus encore le point-virgule et les guillemets sont des signes peu employés et marquant un soin apporté par l'auteur à la langue qu'il emploie ainsi qu'à son lecteur : mettant en exergue un mot, les guillemets peuvent signifier *je sais que ce mot n'est pas tout à fait adapté à la situation mais je n'en trouve pas d'autre*, plaçant délibérément le lecteur en position de juge de la forme d'écriture et du choix des mots ; le point-virgule, signe de ponctuation en voie de disparition, correspond également à un usage "lettré" de la langue.

En identifiant tous les signes de ponctuation présents dans un texte et en les valuant en fonction de leur portée explicitée ici, nous obtenons un indice indiquant dans quelle proportion et selon quelle significativité le texte est ponctué.

Sonorités

Pour chaque texte, nous relevons le nombre de sonorités dont la différence à la normale est significative. Tous le monde connaît *les sanglots longs des violons de l'automne* ou le très fameux *Pour qui sont ces serpents qui sifflent sur nos têtes*. Chacun s'accorde pour y voir une intention de l'auteur : expression de la langueur chez Verlaine, recours à l'onomatopée suggérant les sifflements ophidiens chez Racine. Outre ces cas typiques, il arrive de repérer, en vers ou en prose, une sonorité qui semble sur-représentée et d'en tirer des hypothèses concernant l'effet produit et la volonté de l'auteur. Rien ne dit parfois que ces sonorités ne correspondent pas en fait à la normale, et que ce n'est pas le sens des mots qui pousse le lecteur à voir, dans la présence de liquides ou de dentales, des effets qui ne sont que le propre de la langue française. Ce qui nous intéresse ici n'est pas le nombre de *l* ou de *r* mais le nombre de sonorités sur - ou sous- représentées par rapport à la répartition moyenne des sons de la langue.

Répétitions lexicales

Nous nous sommes également intéressés aux répétitions d'expressions. Si la répétition d'un mot au cours du texte, parfois inévitable, ne crée aucun effet si ce n'est parfois la lassitude du lecteur, la présence répétée d'une suite de mots identiques rythme le texte et participe, au niveau du sens mais plus encore au niveau de la forme, à sa cohérence. Dans *La Nuit de Mai*, Musset joue de cet effet en attribuant à sa Muse des tirades aux débuts identiques : *Poète, prends ton luth et me donne un baiser*, commence-t-elle, pour enchaîner sur *Poète, prends ton luth ; la nuit, sur la pelouse. Ayant entamé sa troisième tirade par Poète, prends ton luth ; le vin de la jeunesse, elle poursuit avec Poète, prends ton luth ; c'est moi, ton immortelle*. L'exemple donné ici est encore une fois extrême : la position de ces propositions en début de tirade les met en évidence et le lecteur ne peut éviter de les remarquer ; l'intention de l'auteur est manifeste. Mais des répétitions lexicales plus discrètes peuvent apparaître : nous les repérons quand la suite lexicale répétée est de trois mots au moins.

Diversité du vocabulaire

Le soin apporté à l'écriture et à la langue se manifestera aussi par la diversité du vocabulaire : dans un même champ lexical, l'auteur a-t-il recours à des termes variés, quasi-synonymes, ou reprend-il fré-

quemment le même mot pour exprimer la même chose ? Pour mesurer ce taux de répétition, nous ne prenons en compte que les mots lexicaux, laissant de côté les mots-outils (déterminants, prépositions, certaines conjonctions...) et les verbes lorsqu'ils sont auxiliaires.

Modalisateurs

La dernière caractéristique prise en compte en ce qui concerne le souci d'écriture est l'utilisation de modalités d'énoncé, dont la fonction est de nuancer le propos : *il semble que, à notre avis, certainement*, sont des mots ou expressions dont l'usage affecte la manière de dire plutôt que le contenu de ce qui est dit. Qu'elles soient justifiées par la crainte de choquer ou la volonté de préciser, les modalités d'énoncé correspondent toujours à une intention de l'auteur sur son lecteur.

CONCLUSION

Les thèmes employés, la variété de construction des phrases, l'organisation des phrases et du texte, le souci d'écriture dont témoigne l'auteur constituent autant d'angles d'analyse d'une production écrite. Travaillant sur des tranches d'âge déterminées, nous obtiendrons rapidement des indices de ce qui ressort des textes produits entre 8 et 11 ans, entre 13 et 15 ans, entre 16 et 17 ans. Mis au regard de l'analyse des processus d'écriture des mêmes textes, les indices relevés donneront une indication de la manière dont, majoritairement, les producteurs arrivent à leur résultat. Il n'est pas question de tirer de ces corrélations des lois incontournables mais plutôt d'observer si l'on peut dégager des relations constantes entre mode d'écriture et type de texte produit. Si l'on ne peut évidemment pas considérer que tel type de texte doit être produit selon tel type de processus et lui seul, la supposition qu'un texte théorique ne s'écrit pas de la même façon qu'un texte narratif est tout à fait recevable. Que cette supposition se confirme ou s'infirme, la quantité d'indices relevés permettra de toute façon des analyses encore inédites de la production écrite à différents moments de l'apprentissage.

Claire DOQUET