



Théorie des langages

LETEILLIER Marion
Numéro d'étudiant : 12400026

UFR LLASIC

Département d'Informatique intégrée en Langues, Lettres et Langage

Master 1 Mention Sciences du Langage Parcours Industries de la Langue

Enseignante responsable : Claude Ponton

Année universitaire
2024 – 2025

Le langage est analysé de plusieurs façons différentes dans la littérature linguistique par de célèbres linguistes. Ferdinand de Saussure, dans *Cours de linguistique générale* en 1916, le décrit comme un système de signes arbitraires qui reposent sur des conventions sociales, et où le langage est une faculté universelle. Noam Chomsky, avec *Aspects of the Theory of Syntax* en 1965, le définit aussi comme une capacité innée du cerveau qui permet de générer des phrases à l'infini à partir de règles de grammaire. André Martinet, lui, dans *Éléments de linguistique générale* en 1960, voit le langage comme un système doublement articulé avec les phonèmes qui s'assemblent en morphèmes. Roman Jakobson, avec *Essais de linguistique générale* en 1963, le définit comme un moyen de communication avec six fonctions principales (référentielle, expressive, conative, etc.). Émile Benveniste considère, dans *Problèmes de linguistique générale* en 1966, que le langage est un système de communication et met en avant son côté subjectif. La perspective du langage est diverse, passant d'un système structuré à un outil de communication. Dans la littérature didactique, Maria Montessori dans *L'esprit absorbant de l'enfant* en 1949, décrit le langage comme une capacité naturelle, universelle qui se développe de manière spontanée chez l'enfant dans un environnement favorable. Le langage, qu'il soit appréhendé par sa seule structure interne ou pour son rôle essentiel dans le développement cognitif et social, est perçu comme naturel. Il offre une multitude d'interprétations possibles à une forme spécifique. À l'inverse, le langage formel lui, n'a qu'une seule interprétation possible. Selon Chomsky, « un langage formel L est formé par un quadruplet : $L = \{V_T, V_N, C, G\}$, où $\{V_T\}$ est un ensemble d'éléments non définis (les mots) appelé *vocabulaire terminal*, $\{V_N\}$ un ensemble d'éléments non définis appelé *vocabulaire non terminal* et comportant un élément P dit *axiome*, C une opération interne sur V_T et sur V_N appelée *concaténation* (ou *juxtaposition de deux éléments*), qui est associative, et G une grammaire formée d'un certain nombre de règles dites *règles de réécriture*. »

Pour répondre au sujet, j'ai constitué un corpus de 10 mémoires d'étudiants déposés dans la plateforme DUMAS. J'ai choisi le domaine des Sciences de l'Homme et de la Société avec une spécification sur le Langage. Correspondant à un thème assez vaste de sujets, cela augmente les chances de trouver des annonces de plan différentes les unes des autres. Cela permet d'éviter les biais liés à la répétition de la même structure pour que l'expression régulière ne se limite pas aux particularités de ce corpus. L'idéal est que cette expression soit appliquée à n'importe quel texte.

Le mémoire 1 traite de l'impact des jeux coopératifs sur la posture d'élève apprenant en maternelle et le mémoire 2 montre comment le jeu peut améliorer les compétences syntaxiques des élèves en maternelle. Le mémoire 3 analyse l'entrée en littérature des élèves UPE2A, quant au mémoire 4, il se concentre sur la modélisation et l'enrichissement sémantique des catalogues de manuscrits. Le mémoire 5 traite de la résistance des enseignants de maternelle face aux changements pédagogiques avec le langage oral, et le mémoire 6 aborde la place du livre de jeunesse dans les pratiques des assistants maternels pour développer les compétences langagières des enfants. Le mémoire 7 s'intéresse aux interactions langagières dans l'enseignement de la numération au cycle 2. Le mémoire 8 montre l'introduction de la division posée au cycle 3, et le mémoire 9 étudie les postures de lecteur au cycle 3 pour améliorer les pratiques de lecture. Enfin, le mémoire 10 compare les troubles vocaux chez les étudiants de différents cursus.

Après avoir élaboré sept expressions régulières distinctes, je les ai combinées en une seule, puis affinées en tenant compte des autres annonces de plan afin de couvrir l'ensemble du corpus. Cette approche a fonctionné, mais pour vérifier si elle est adaptée à n'importe quelle annonce de plan, je l'ai testée sur un mémoire répondant aux mêmes critères de sélection (DUMAS, Sciences de l'Homme et de la Société, Langage) ainsi que sur un autre dans une discipline différente.

Voici mes expressions régulières spécifiques à une annonce de plan :

- \bDans un premier temps\b,?\s*(.*?\.)\s*\bEnsuite\b,?\s*(.*?\.)\s*\bEnfin\b,?\s*(.*?\.)\s*\$
- \bDans une première partie\b,?\s*(.*?\.)\s*\bNous rappellerons\b,?\s*(.*?\.)\s*\bDans une seconde partie\b,?\s*(.*?\.)\s*\bDans cette même partie\b,?\s*(.*?\.)\s*\bEnfin dans une troisième et dernière partie\b,?\s*(.*?\.)\s*\$
- \bLa première partie\b (est consacrée.*?\.)\s*(Elle permet .*?\.)\s*\bLa seconde partie\b (s'attache .*?\.)\s*(Cet alignement .*?\.)\s*\bEnfin\b,?\s*\bla troisième partie\b (s'attache .*?\.)\s*\$
- \bDans une première partie\b,?\s*(.*?\.)\s*\bDans une deuxième partie\b,?\s*(.*?\.)\s*\bDans une troisième partie\b,?\s*(.*?\.)\s*\bEnfin\b,?\s*(la dernière partie .*?\.)\s*\$
- \bPour trouver des éléments de réponse à ce questionnement\b,?\s*(.*?\.)\s*\bNous développerons ensuite\b,?\s*(.*?\.)\s*\bUne troisième partie\b,?\s*(.*?\.)\s*\bLe résultat\b (de ce travail sera .*?\.)\s*\$
- \bDans une première partie\b,?\s*(.*?\.)\s*\bPar ailleurs\b,?\s*(.*?\.)\s*\bDans une deuxième partie\b,?\s*(.*?\.)\s*\bPuis\b,?\s*(nous analyserons.*?\.)\s*\$
- \bEn vue de répondre à cette question de recherche\b,?\s*(.*?\.)\s*\bEnsuite\b,?\s*(.*?\.)\s*\bUne partie\b (est exclusivement dédiée .*?\.)\s*\bFinalement\b,?\s*(s'en suit .*?\.)\s*\bUne seconde partie\b,?\s*(.*?\.)\s*\bSuite à cela\b,?\s*(.*?\.)\s*\bFinalement\b,?\s*(nous présenterons .*?\.)\s*\$

Voici les expressions régulières fusionnées en une seule, ajustée pour intégrer les variations des autres annonces de plan :

- \b(Dans un premier temps|dans un premier temps|Dans une première partie|Pour trouver des éléments de réponse à ce questionnement|En vue de répondre à cette question de recherche|La première partie|Nous aborderons)\b,?\s*(.*?\.)\s*\b(Ensuite|Puis|Dans une deuxième partie|Dans une seconde partie|Nous développerons ensuite|Par ailleurs|Suite à cela|Le résultat|Une seconde partie|La seconde partie)\b,?\s*(.*?\.)\s*\b(Dans une troisième partie|dans une troisième et dernière partie|En conclusion|Pour conclure|Enfin|Puis|Finalement|Une partie|Une troisième partie|dans une troisième section)\b,?\s*(.*?\.)\s*(\b(Finalement|Enfin)|\b,?\s*(.*?\.)|La dernière partie\s+.*?\.)?

J'ai modifié l'expression régulière afin de la rendre insensible à la casse avec (?i) et de la rendre plus générale :

- (?i)\b(Dans un premier temps|Dans une première partie|La première partie||Une première partie)\b,?\s*(.*?\.)\s*\b(Ensuite|Puis|Dans une deuxième partie|Dans une seconde partie|Par ailleurs|Suite à cela|Une seconde partie|La seconde partie)\b,?\s*(.*?\.)\s*\b(Dans une troisième partie|Enfin|Puis|Finalement|Une partie|Une troisième partie|Dans une troisième section)\b,?\s*(.*?\.)\s*(\b(Finalement|Enfin|Dans une dernière partie)\b,?\s*(.*?\.)|La dernière partie\s+.*?\.)?

En analysant les annonces de plan du corpus, plusieurs problèmes se posent concernant la généralisation avec une expression régulière. Tout d'abord, les annonces de plan ne suivent pas un schéma purement strict comme « Dans un premier (...), dans un second temps (...), dans un troisième temps (...) ». Certaines annonces peuvent commencer par cette structure, mais continuent avec d'autres

connecteurs, comme « Puis (...) Enfin (...) » (annonce de plan 1). D'autres encore procèdent à une alternance avec « La première partie (...) La seconde partie (...) Enfin (...) » (annonce de plan 4). Même lorsque la structure semble linéaire, des éléments viennent ajouter de la complexité à l'analyse. Dans l'annonce de plan 3, il y a un ajout avec « Enfin dans une troisième et dernière partie (...) », ce qui peut perturber la détection d'annonce de plan avec du bruit. De plus, le nombre de parties varie souvent. La structure en trois parties est la plus courante, mais certaines incluent une quatrième ou une cinquième partie, comme dans les annonces de plan 1, 5, 6, 8 et 9. Par ailleurs, des annonces se structurent seulement avec des connecteurs logiques comme « Ainsi (...) Puis (...) Enfin (...) », sans introduction explicite d'une première ou seconde partie (annonce de plan 1). À l'inverse, d'autres annonces ne commencent pas par des connecteurs ou des phrases d'introduction, les rendant plus implicites. Ce phénomène entraîne un silence pour l'expression régulière, qui ne détecte pas la structure. Il y a aussi le cas des annonces de plan qui ont leurs connecteurs ou phrases introductives en milieu de phrase, et qui ne sont pas détectées en raison de l'absence de majuscule (sensible à la casse).

Tout cela montre qu'une standardisation s'avère compliquée au-delà du corpus et que les annonces de plan ne sont pas régulières. Le style ou la structure des mémoires ne sont pas semblables selon les disciplines : souvent plus rigides en mathématiques, plus libres et variées en littérature ou sciences humaines. L'expression régulière n'a pas fonctionné sur l'annonce de plan d'un autre mémoire de Science de l'Homme et de la Société portant sur le Langage (test1), mais a réussi sur un mémoire de Chimie (test2). Cela illustre qu'il est impossible de généraliser un langage formel pour analyser un langage naturel. Les variations interindividuelles propres au langage naturel mettent en avant des différences en termes de structure et de style, qui échappent aux règles du langage formel. L'annonce de plan est un phénomène complexe, car linguistiquement, elle est réalisée sous plein de manières différentes. Même s'il est possible d'écrire un formalisme à l'aide d'une expression régulière qui décrirait ce phénomène de manière optimale (très courant en TAL), ce formalisme ne sera pas précis à 100 %.

Corpus

[Annonce plan 1 : L'impact des jeux coopératifs sur l'acquisition de la posture d'élève apprenant à l'école maternelle - DUMAS - Dépôt Universitaire de Mémoires Après Soutenance](#)

Ainsi, dans le cadre de ce mémoire, nous présenterons dans un premier temps l'état de l'art, qui abordera plus en détail l'importance du jeu à l'école maternelle, la coopération à l'école et la définition de la posture d'élève apprenant en s'appuyant sur deux axes fondamentaux : le langage et l'attention en classe. Puis, nous présenterons notre problématique ainsi que nos hypothèses, suivies de la méthode qui a été mise en place dans le cadre de cette étude. Enfin, nous aborderons les résultats de cette recherche et une partie discussion qui analysera ces résultats afin de les mettre en lien avec la problématique et les hypothèses de cette étude, avant de conclure.

[Annonce plan 2 : Améliorer les compétences syntaxiques grâce au jeu - DUMAS - Dépôt Universitaire de Mémoires Après Soutenance](#)

Dans un premier temps, nous présenterons une partie théorique en mettant en avant les différentes étapes d'acquisition du langage ainsi que l'impact de l'environnement sur l'apprentissage de l'oral, avant d'explorer les différentes méthodes pour travailler la syntaxe en maternelle. Ensuite, nous exposerons le dispositif mis en place lors de notre séquence pédagogique tout en décrivant et justifiant notre

méthode utilisée. Enfin, nous présenterons nos résultats qui mettent en évidence le développement de compétences syntaxiques chez les élèves, à la suite de la mise en place de cette séquence.

Dans un premier temps Ensuite **Enfin**

Annonce plan 3 : L'entrée en littérature en UPE2A : un pari impossible ? Proposition d'un premier lexique pour développer la conscience phonologique et initier à la grammaire (FLS) - DUMAS - Dépôt Universitaire de Mémoires Après Soutenance

Dans une première partie, nous évoquerons les théories concernant l'entrée dans l'écrit en Français langue maternelle (FLM) puis en Français langue seconde (FLS), ainsi que les différences à prendre en compte. Nous rappellerons également la politique de scolarisation des EANA. Dans une seconde partie, nous présenterons la méthodologie, le contexte de notre recherche ainsi que le profil de nos sujets d'étude. Nous évoquerons notre enquête auprès des enseignants de FLS/FLSco en charge des dispositifs UPE2A de l'Académie Nancy-Metz. Dans cette même partie, nous proposerons un corpus de structures syntaxiques à enseigner aux apprenants grands débutants, basé notamment sur leur identité. Enfin dans une troisième et dernière partie, nous rendrons compte des apports de ce premier lexique sur les apprentis-lecteurs en FLS et les approches de la grammaire française pouvant en découler.

Dans une première partie Nous rappellerons **Dans une seconde partie** **Dans cette même partie** **Enfin dans une troisième et dernière partie**

Annonce plan 4 : Modélisation, enrichissement sémantique et diffusion d'un corpus textuel semi-structuré : le cas des catalogues de vente de manuscrits - DUMAS - Dépôt Universitaire de Mémoires Après Soutenance

La première partie est consacrée à la production et à la modélisation des données. Elle permet de mieux comprendre la nature et la structure des documents traités, et la manière dont celle-ci peut être mise à profit du traitement automatisé des documents. La seconde partie s'attache à présenter une chaîne de traitement pour l'alignement automatisé des entrées de catalogues avec Wikidata. Cet alignement s'appuie entièrement sur des méthodes de détection de motifs, et donc sur la nature semi-structurée des documents. Cette chaîne de traitement vise à produire des données qui permettront d'étudier l'impact de facteurs biographiques sur le prix d'un manuscrit. Ce processus d'alignement est également l'occasion de discuter des méthodes mises en place dans les humanités numériques centrées sur le texte pour le traitement automatisé du langage. Enfin, la troisième partie s'attache à présenter les objectifs, le fonctionnement et les questionnements soulevés par le développement de KatAPI.

La première partie (est consacrée.) (Elle permet .) **La seconde partie** (s'attache .) (Cet alignement .) **Enfin** **la troisième partie** (s'attache .)

Annonce plan 5 : La résistance au changement chez les enseignants de maternelle face aux guides d'accompagnement des programmes de l'école maternelle dans le domaine du langage oral - DUMAS - Dépôt Universitaire de Mémoires Après Soutenance

Dans une première partie, nous présenterons les données théoriques liées à l'acquisition et l'enseignement du langage oral et du lexique, celles liées à la résistance au changement et notamment la résistance au changement des enseignants. Nous préciserons notre questionnement initial à partir

d'une problématique et nous expliciterons nos hypothèses de travail. Dans une deuxième partie, nous présenterons le protocole expérimental en lien avec la résistance au changement des enseignants de maternelle face aux guides d'accompagnement des programmes dans le domaine du langage oral. Nous exposerons le contexte et justifierons les choix méthodologiques retenus pour recueillir et analyser les résultats de notre recherche. Dans une troisième partie, nous analyserons les données recueillies au travers d'un questionnaire auto-déclaratif, d'entretiens semi-directifs et d'enregistrements de séances de langage oral. Enfin, la dernière partie nous amènera à discuter des résultats et à valider ou invalider les hypothèses puis à dégager les limites et les perspectives de cette recherche avant de conclure par les nouvelles questions qu'elle engendre.

\bDans une première partie\b,?\s(.*?\.)\s*\bDans une deuxième partie\b,?\s*(.*?\.)\s*\bDans une troisième partie\b,?\s*(.*?\.)\s*\bEnfin\b,?\s*(la dernière partie .*\?)\s\$*

Annonce plan 6 : La place du livre de jeunesse dans les pratiques professionnelles des assistants maternels - DUMAS - Dépôt Universitaire de Mémoires Après Soutenance

Pour trouver des éléments de réponse à ce questionnement, nous explorerons les ouvrages scientifiques pour accroître nos connaissances sur la littérature de jeunesse et sur son impact sur le développement des compétences langagières chez l'enfant. Nous développerons ensuite nos connaissances sur le public étudié en définissant le métier d'assistant maternel et sur leur formation. Une troisième partie permettra de définir le protocole de recherche mis en place en commençant par définir la problématique et les hypothèses de recherche formulés. Le résultat de ce travail sera développé dans une dernière partie.

\bPour trouver des éléments de réponse à ce questionnement\b,?\s(.*?\.)\s*\bNous développerons ensuite\b,?\s*(.*?\.)\s*\bUne troisième partie\b,?\s*(.*?\.)\s*\bLe résultat\b (de ce travail sera .*\?)\s\$*

Annonce plan 7 : Développer les interactions langagières dans l'enseignement et l'apprentissage de la numération au cycle 2 - DUMAS - Dépôt Universitaire de Mémoires Après Soutenance

Dans une première partie, nous allons définir la numération, ses enjeux et son langage associé. Par ailleurs, l'élaboration de la problématique fait ressortir l'importance de questionner le choix variables didactiques au regard de l'apprentissage de la numération et de son langage au cycle 2. Dans une deuxième partie nous expliquerons la démarche que nous avons utilisée lors de c 7 recherche. Puis nous analyserons les résultats dans le but de répondre à notre question de recherche et valider ou non nos hypothèses.

\bDans une première partie\b,?\s(.*?\.)\s*\bPar ailleurs\b,?\s*(.*?\.)\s*\bDans une deuxième partie\b,?\s*(.*?\.)\s*\bPuis\b\s*(nous analyserons.*?\.)\s\$*

Annonce plan 8 : Introduction de la division posée au cycle 3 : articulation entre langage, manipulation et algorithme de la division posée - DUMAS - Dépôt Universitaire de Mémoires Après Soutenance

En vue de répondre à cette question de recherche, une première partie consacrée à l'enseignement du concept et de l'algorithme de la division posée permet d'appréhender les évolutions qu'a subi l'opération. En conséquence, l'étude d'un ancien algorithme ainsi que de celui que nous utilisons actuellement en France en témoigne. Dans une section dédiée, ces deux algorithmes sont confrontés grâce à l'identification de critères de comparaisons. Ensuite, nous verrons plus spécifiquement les divisions rencontrées par les élèves de l'école élémentaire puis le langage utilisé lors de la pose de

l'opération. Une partie est exclusivement dédiée à la contextualisation de la division au travers des situations et activités de partages. L'intérêt porté aux difficultés des élèves permet de nous éclairer sur les résistances auxquelles ils se confrontent lors de l'application de l'algorithme de la division posée. L'enquête CEDRE (Cycle des Évaluations Disciplinaires Réalisées sur Échantillons) constitue un apport majeur. Finalement, s'en suit l'annonce de la question de recherche ainsi que des hypothèses émises. Une seconde partie dédiée à la méthode de notre recherche permet d'avancer les choix effectués pour réaliser le recueil de données. Dans le même temps, une attention est portée aux analyses a priori et a posteriori de l'expérimentation. Suite à cela, dans une troisième section, nous aurons l'occasion d'apprécier les résultats ainsi que de les discuter au regard de la littérature. Finalement, nous présenterons les intérêts et limites de notre recherche avant de conclure.

\bEn vue de répondre à cette question de recherche\b,?\s(.*?\.)\s*\bEnsuite\b,?\s*(.*?\.)\s*\bUne partie\b (est exclusivement dédiée .*?\.)\s*\bFinalement\b,?\s*(s'en suit .*?\.)\s*\bUne seconde partie\b,?\s*(.*?\.)\s*\bSuite à cela\b,?\s*(.*?\.)\s*\bFinalement\b,?\s*(nous présenterons .*?\.)\b*

Annonce plan 9 : Les postures de lecteur au cycle 3 : les rendre explicites pour faire évoluer le rapport des élèves à la lecture - DUMAS - Dépôt Universitaire de Mémoires Après Soutenance

Je proposerai dans un premier temps une synthèse des travaux, réflexions, et textes officiels relatifs au sujet lecteur et à l'enseignement de la littérature, avant de définir la problématique de cette étude. Dans une deuxième partie, je décrirai le protocole expérimental, avant de présenter les résultats dans une troisième partie. Enfin, les résultats de l'expérimentation seront analysés et discutés dans la dernière partie de ce mémoire.

Annonce plan 10 : Comparaison du voice handicap index (VHI) d'étudiants niveau master en Picardie. Population cible : étudiants en master métiers de l'enseignement, de l'éducation et de la formation (MEEF) et des étudiants niveau master en maïeutique, médecine et psychologie - DUMAS - Dépôt Universitaire de Mémoires Après Soutenance

Nous aborderons des rappels anatomo-physiologiques, les paramètres acoustiques de la voix, la diversité des manifestations vocales. Puis, nous présenterons les troubles de la voix, en explicitant la notion de charge vocale et la dysphonie dysfonctionnelle. Ensuite, nous aborderons les facteurs de risque, en commençant par les facteurs de risques internes à tous les individus et ceux spécifiques à l'enseignement. Dans un second temps, nous développerons notre problématique et nos hypothèses. Puis nous présenterons la méthodologie utilisée dans notre étude. Nous analyserons ensuite nos résultats statistiques et nous les mettrons en contraste avec nos hypothèses de départ

Test 1 – Mémoire Science de l'Homme et de la Société – Langage L'inclusion scolaire d'élèves déficients auditifs – DUMAS – Dépôt Universitaire de Mémoires Après Soutenance

Dans un premier temps, je me suis appuyée sur des textes scientifiques pour apporter des informations sur la communication des élèves déficients auditifs à l'école, tout en effectuant un rappel historique sur l'inclusion scolaire, avant de présenter mon contexte d'exercice. Dans un second temps, je présenterai les adaptations didactiques et pédagogiques, où une réflexion autour d'outils a été faite avant d'effectuer les mises en œuvre au sein de ma classe, puis je présenterai les résultats obtenus et l'analyse de mes mises en œuvre.

Test 2 – Mémoire Chimie Mise en place et évaluation des performances d'une nouvelle gamme de système de chromatographie liquide ultra-performance (UPLC) - DUMAS - Dépôt Universitaire de Mémoires Après Soutenance

Dans une première partie, nous verrons brièvement le fonctionnement de l'HPLC et des notions fondamentales liées à la chromatographie liquide. Puis nous verrons l'apport de la technologie UPLC vis-à-vis de l'HPLC avant de faire un bref rappel du fonctionnement d'un détecteur UV et MS avec un analyseur quadripolaire. Ensuite dans une seconde partie, nous présenterons les protocoles utilisés au cours de ce stage afin de réaliser les quatre études utilisées. Puis nous évaluerons les différences et les similitudes entre les instruments proposés par Waters. Enfin dans une dernière partie, après avoir réalisé un traitement statistique des résultats, nous déterminerons la performance des instruments UPLC de gamme Acquity, des détecteurs UV de la gamme Acquity et les spectromètres de masse de la gamme quadripolaire.

RegExr: Learn, Build, & Test RegEx

Character classes

. any character except newline
\w\d\s word, digit, whitespace
\W\D\S not word, digit, whitespace
[abc] any of a, b, or c
[^abc] not a, b, or c
[a-g] character between a & g

Anchors

^abc\$ start / end of the string
\b\B word, not-word boundary

Escaped characters

\. * \\ escaped special characters
\t \n \r tab, linefeed, carriage return

Groups & Lookaround

(abc) capture group
\1 backreference to group #1
(?:abc) non-capturing group
(?=abc) positive lookahead
(?!abc) negative lookahead

Quantifiers & Alternation

a*a+a? 0 or more, 1 or more, 0 or 1
a{5}a{2,} exactly five, two or more
a{1,3} between one & three
a+?a{2,}? match as few as possible
ab|cd match ab or cd