

# LA PRÉDICTION LINGUISTIQUE EN FRANÇAIS : DE LA THÉORIE À LA PRATIQUE

## Linguistic Prediction in French: From Theory to Practice

---

### Mathieu GUIDERE

Professeur à l'Université de Paris  
Directeur de recherches (INSERM, France)  
[mathieu.guidere@u-paris.fr](mailto:mathieu.guidere@u-paris.fr)

### Lahcen ELGHAZI

Professeur à l'Université du Québec  
à Montréal (UQÀM, Canada)  
[elghazi.lahcen@uqam.ca](mailto:elghazi.lahcen@uqam.ca)

---

**Résumé :** Cet article pose les jalons d'une grammaire prédictive du français, en s'inscrivant dans la continuité des travaux récents sur la linguistique prédictive. Partant de l'idée que la compréhension et la production langagières mobilisent en permanence des mécanismes d'anticipation cognitive, il interroge les caractéristiques structurelles du français qui en font un objet d'analyse prédictive particulièrement féconde : la richesse de son orthographe morphologique, l'opposition entre graphie et phonie, la densité de son appareil déterminatif et la codification serrée de ses usages sociolinguistiques. L'article déploie ensuite une grammaire prédictive organisée selon trois ordres d'analyse. Au niveau sémantique, il met en lumière la manière dont les solidarités lexicales, les préfixes et suffixes productifs ainsi que les locutions figées contribuent à baliser l'anticipation de l'unité lexicale suivante. Au niveau syntaxique, il examine le rôle des pronoms clitiques, des connecteurs logiques, des constructions corrélatives et de la cohérence des temps verbaux dans la projection de la phrase à venir. Au niveau pragmatique, il analyse les rituels de civilité, les scénarios d'invitation et de repas, les stratégies d'atténuation et les structures rhétoriques qui organisent la prévisibilité des échanges en français. L'article se conclut sur les enjeux d'une telle approche pour la conception de modèles de langage prédictifs en intelligence artificielle, en défendant l'idée d'architectures linguistiquement fines et culturellement calibrées, aptes à rendre compte des singularités typologiques et discursives propres à chaque langue.

**Mots-clés :** langue française, grammaire prédictive, prédiction sémantique, prédiction syntaxique, prédiction pragmatique, didactique des langues, modèles de langage, intelligence artificielle, adaptation culturelle, rhétorique.

---

**Abstract:** This article sets out the foundations of a predictive grammar of French, in line with recent developments in predictive linguistics. Taking as its starting point the idea that linguistic comprehension and production constantly mobilise cognitive anticipation mechanisms, it investigates the structural features of French that make it a particularly productive object of predictive analysis: the richness of its morphological orthography, the gap between writing and pronunciation, the density of its determiner system, and the tight codification of its sociolinguistic usages. The article then unfolds a predictive grammar organised along three analytical axes. At the semantic level, it highlights how lexical solidarities, productive prefixes and suffixes, and frozen expressions contribute to guiding the anticipation of the next word. At the syntactic level, it examines the role of clitic pronouns, logical connectors, correlative constructions, and the coherence of verb tenses in projecting the upcoming sentence. At the pragmatic level, it analyses the rituals of civility, the scripts of invitation and meal-sharing, mitigation strategies, and rhetorical structure that organise the predictability of exchanges in French. The article closes with the implications of such an approach for the design of predictive language models in artificial intelligence, arguing for linguistically refined and culturally calibrated architectures capable of reflecting the typological and discursive specificities proper to each language.

**Keywords:** French language, predictive grammar, semantic prediction, syntactic prediction, pragmatic prediction, language teaching, language models, artificial intelligence, cultural adaptation, rhetorics.

---

## Introduction

Parmi les renouvellements théoriques qui traversent la linguistique contemporaine, la question de la prédiction occupe une position stratégique. Il s'agit d'éclairer la capacité du système cognitif à formuler en continu, au fil du traitement du signal langagier, des hypothèses sur les unités à venir, à partir d'un faisceau d'indices phonétiques, morphologiques, lexicaux, syntaxiques et discursifs. Cette perspective rompt avec le modèle linéaire de décodage qui supposait un traitement strictement séquentiel : comprendre un énoncé revient à projeter, réviser et stabiliser des représentations anticipées, et non à reconstituer après coup une chaîne sonore ou graphique.

Si ce mécanisme d'anticipation relève d'une architecture cognitive universelle, sa mise en œuvre effective varie d'une langue à l'autre. Les recherches pionnières dans le domaine se sont principalement appuyées sur des corpus anglophones, ce qui a longtemps orienté la modélisation théorique. L'élargissement du cadre à des systèmes linguistiques distincts fait apparaître que les ressorts de la prédiction tiennent étroitement à des paramètres typologiques précis : degré de flexion morphologique, liberté de l'ordre des mots, richesse du marquage grammatical, vigueur des conventions socio-discursives. C'est dans cet horizon comparatiste que se situe le présent travail, consacré cette fois au français.

Le français offre, du point de vue de la linguistique prédictive, un profil singulier. Langue romane issue d'une évolution intense du latin populaire, elle a connu une érosion notable de sa flexion orale tout en conservant, à l'écrit, un système morphologique dense et en partie silencieux. Cette dissymétrie entre code oral et code écrit fait du français un cas d'étude précieux : à l'oral, la prédiction s'appuie largement sur les déterminants, les pronoms clitiques et les marques prosodiques, là où à l'écrit elle exploite un riche réseau de désinences, d'accords et de terminaisons muettes. Loin d'être un appauvrissement, cette double économie constitue une ressource prédictive originale.

À ces traits structurels s'ajoute une tradition normative et rhétorique d'une densité remarquable. Le français a fait l'objet, depuis le XVII<sup>e</sup> siècle au moins, d'un travail de codification explicite qui a modelé l'usage et cristallisé des conventions discursives particulièrement stables : formules de politesse, rituels épistolaires, scénarios de civilité et

d'hospitalité, clôture des échanges, gestion des registres. Ces stabilités, loin d'être marginales, constituent des points d'ancrage essentiels pour toute modélisation prédictive de la langue.

Le présent article se propose d'esquisser les fondements d'une grammaire prédictive du français, pensée à la fois comme outil d'analyse et comme instrument didactique au service de l'apprentissage de la langue. Cette grammaire s'articule autour de trois niveaux d'anticipation : le niveau sémantique, où se joue la probabilité du mot à venir ; le niveau syntaxique, où se dessinent les structures de phrase attendues ; et le niveau pragmatique, où se déploient les scénarios conversationnels et narratifs. Ces trois plans, on le verra, ne fonctionnent pas en parallèle, mais en étroite interaction.

## La spécificité prédictive de la langue française

Le français présente un ensemble de propriétés qui, prises dans leur entrelacement, définissent un profil prédictif propre. Ce profil ne se réduit ni à celui d'une langue fortement flexionnelle comme le russe ou l'allemand, ni à celui d'une langue analytique comme l'anglais : il occupe une position intermédiaire qui mérite une description spécifique.

Un premier trait concerne la dissociation entre oral et écrit. Dans la chaîne parlée, un grand nombre de marques grammaticales ne se réalisent pas phonétiquement : les désinences verbales du pluriel de la première conjugaison (*ils parlent / il parle*), les marques du pluriel nominal (*le livre / les livres* sont partiellement distingués par le seul déterminant) ou certaines terminaisons adjectivales s'effacent à l'oral. Cette discrétion phonique confère au déterminant et au pronom clitique un rôle prédictif majeur : ce sont eux qui portent, en amont du nom ou du verbe, l'information grammaticale qui orienterait autrement l'anticipation. Dès qu'un locuteur francophone entend *les*, le système cognitif active massivement les noms pluralisables et ajuste les représentations en conséquence, avant même que le nom ne soit produit.

Un deuxième trait tient à la densité du paradigme déterminatif. Le français oppose avec précision articles définis, indéfinis, partitifs, démonstratifs et possessifs, chaque série porteuse d'informations prédictives spécifiques. L'article partitif *du, de la, des* signale par exemple une saisie non comptable d'une entité et projette typiquement un nom massif ou un substantif de consommation : *boire du vin, manger de*

*la soupe*. L'opposition *ce livre / ce livre-ci / ce livre-là* introduit une dimension déictique qui oriente l'anticipation vers un référent situé dans l'espace ou dans le discours. Ces micro-variations forment un réseau d'indices à forte valeur prédictive.

Un troisième trait, d'ordre pragmatique et culturel, concerne la vitalité des conventions de politesse et des rituels sociolinguistiques. L'opposition *tu / vous*, l'obligation quasi systématique du *bonjour* en ouverture d'interaction, la pratique codifiée du compliment gastronomique, la fréquence des tournures d'atténuation (*je me permets de, si cela ne vous dérange pas, auriez-vous l'obligeance de*) façonnent un paysage où la maîtrise du prévisible vaut autant, sinon davantage, que celle des règles grammaticales. L'héritage classique, puis le réinvestissement contemporain de ces codes, construisent une prévisibilité socio-culturelle que toute grammaire prédictive du français se doit d'intégrer.

Ces trois traits (dissociation oral/écrit, densité du système déterminatif, codification pragmatique des usages) ne sont pas des accidents de surface : ils dessinent ensemble l'ossature prédictive du français et appellent une description dédiée.

### Grammaire prédictive du français

La grammaire prédictive du français se démarque des approches normatives traditionnelles par un changement de visée. Il ne s'agit plus de recenser des règles prescriptives ni de dresser un inventaire exhaustif de paradigmes, mais de décrire les régularités probabilistes qui guident l'enchaînement des unités langagières en situation. Dans cette optique, la dichotomie classique entre l'« acceptable » et l'« inacceptable » laisse place à une échelle continue allant des configurations hautement attendues aux configurations marginales, en passant par des zones de plausibilité moyenne.

Ce déplacement modifie en profondeur la notion même de compétence linguistique. Apprendre le français ne revient plus uniquement à mémoriser des tableaux de conjugaison ou à appliquer des règles d'accord, mais à cultiver une sensibilité fine aux configurations les plus probables dans un contexte donné. Cette compétence se déploie sur trois plans interdépendants, qui structurent les sections qui suivent.

### La prédiction sémantique en français

La prédiction sémantique désigne l'anticipation du lexème le plus vraisemblable à un point donné de l'énoncé. En français, elle mobilise trois leviers principaux : les solidarités lexicales qui organisent les champs thématiques, les contraintes morphologiques portées par les préfixes et suffixes productifs, et les expressions figées qui cristallisent des continuations quasi obligatoires.

Le premier levier repose sur les réseaux de cohésion lexicale. Dès qu'un champ sémantique est ouvert, l'ensemble des termes qui s'y rattachent gagne en accessibilité cognitive. La productivité dérivationnelle du français amplifie ce phénomène : à partir d'une base comme *école*, le locuteur projette aisément *écolier, écolière, scolaire, scolarité, scolariser, scolarisation*. La reconnaissance de la racine déclenche l'activation d'une constellation morphologique et sémantique qui guide l'anticipation.

Le deuxième levier concerne les affixes à valeur catégorielle. Le préfixe *re-* signale régulièrement la répétition ou le retour : *refaire, relire, revenir, rejouer*. Le suffixe *-eur / -euse* construit des noms d'agent ou d'instrument : *chanteur, chanteuse, ordinateur, aspirateur*. Le suffixe *-tion* marque un procès ou son résultat : *formation, rédaction, livraison*. Le suffixe *-able / -ible* exprime la possibilité passive : *lisible, mangeable, admissible*. Dès que l'un de ces affixes est perçu, l'interprétation catégorielle et sémantique du mot se stabilise, y compris lorsqu'il s'agit d'un néologisme.

Le troisième levier est celui des collocations et des locutions figées, dont l'amorce appelle presque mécaniquement sa suite. Ainsi, *pousser un \_\_* dans un contexte d'émotion vive appelle *cri* ou *soupir* ; *rendre un \_\_* dans un contexte scolaire ou professionnel appelle *devoir* ou *service* ; *tirer des \_\_* dans un contexte réflexif appelle *conclusions* ou *leçons*. Les expressions proverbiales fonctionnent selon la même logique : *mieux vaut \_\_* appelle *tard que jamais* ou *prévenir que guérir*.

**Exemple 1 — Champ lexical thématique :** En contexte universitaire, l'apparition du verbe *s'inscrire* active un espace sémantique très restreint et projette des compléments tels que *en licence, à la faculté, dans un cursus, à l'examen*. Un complément étranger à ce champ, bien que grammaticalement recevable, serait statistiquement très peu probable.

**Exemple 2 — Affixe et catégorie :** La perception du suffixe *-iste* en contexte évaluatif ou professionnel oriente l'attente vers un nom d'adepte ou de praticien : *un journaliste chevronné, une dentiste consciencieuse, un pianiste virtuose*. Le suffixe opère comme un marqueur catégoriel prédictif robuste.

**Exemple 3 — Polarité argumentative :** L'énoncé *Il n'est pas seulement \_\_, il est aussi \_\_* contraint les deux vides à s'inscrire dans la même polarité. Si le premier est mélioratif (*compétent*), le second doit l'être également (*bienveillant, disponible, rigoureux*). Une rupture de polarité (*Il n'est pas seulement compétent, il est aussi fatigué*) serait perçue comme incohérente, sauf effet stylistique recherché.

**Exemple 4 — Prédiction par contraste concessif :** Dans *En dépit du froid glacial, \_\_, la structure concessive projette une proposition qui marque la persévérance ou un résultat contre toute attente : nous avons tout de même fait la randonnée, la cérémonie s'est tenue à l'extérieur, les spectateurs sont restés jusqu'à la fin*.

**Exemple 5 — Locutions cristallisées :** L'amorce *coûte que \_\_* appelle *coûte* ; *bon an \_\_* appelle *mal an* ; *vaille que \_\_* appelle *vaille* ; *à tes \_\_*, dans un contexte d'éternuement, appelle *souhaits*. Ces configurations atteignent un degré de prédictibilité quasi maximal, au point que l'interruption de la formule figée produit un effet stylistique marqué.

**Exemple 6 — Collocation culturelle :** En contexte gastronomique, *accorder les \_\_* projette *mets et les vins* ; en contexte politique, *faire passer la \_\_* projette *réforme* ou *loi* ; en contexte scolaire, *passer un \_\_* projette *examen* ou *concours*. La cristallisation collocationnelle dépend du registre discursif activé.

## La prédiction syntaxique en français

La prédiction syntaxique porte sur l'anticipation des structures phrastiques et des relations de dépendance entre constituants. Le français dispose de plusieurs dispositifs dont la valeur prédictive est particulièrement élevée : les pronoms clitiques antéposés au verbe, les connecteurs logiques et conjonctions de subordination, les constructions corrélatives, la concordance des temps, et la négation en deux éléments.

Le premier ressort réside dans la valence verbale et dans le système prépositionnel qui l'accompagne. Le choix du verbe conditionne fortement la suite

syntactique : *penser à* contraint un complément introduit par *à*, *rêver de* impose un complément en *de*, *obliger quelqu'un à* appelle un infinitif, *empêcher quelqu'un de* en appelle un autre. De nombreux verbes construisent leurs compléments selon des schémas strictement codifiés (*se souvenir de, se rendre compte de, veiller à, tenir à*) dont la reconnaissance oriente immédiatement la suite de l'énoncé.

Le deuxième ressort tient aux pronoms clitiques, qui précèdent le verbe en français et en anticipent la forme argumentale. La séquence *Je le lui ai \_\_* projette un participe passé dont les accords sont déjà partiellement contraints, et appelle un verbe à double complémentation (*donné, remis, envoyé, expliqué*). De même, l'apparition du pronom *en* projette typiquement un verbe compatible avec un complément en *de* ou un partitif : *Des livres, j'en ai \_\_*, qui appelle soit un participe (*lus, vus, achetés*), soit un quantificateur (*plusieurs, beaucoup, trois*).

Le troisième ressort relève des connecteurs logiques et des conjonctions de subordination, qui opèrent comme de véritables opérateurs de branchement syntaxique. *Bien que* et *quoique* imposent le subjonctif : *Bien qu'il soit épuisé, il poursuit son travail*. *Pour que, afin que, de sorte que* (finalité) appellent également le subjonctif : *Je t'explique cela pour que tu comprennes la situation*. *Puisque, étant donné que, vu que* annoncent en revanche une cause à l'indicatif et projettent une principale argumentativement cohérente. *Si*, enfin, sélectionne un schéma temporel strict : *si* + présent → futur simple ; *si* + imparfait → conditionnel présent ; *si* + plus-que-parfait → conditionnel passé.

Le quatrième ressort concerne les constructions corrélatives et clivées. Les corrélations *non seulement... mais aussi..., soit... soit..., d'une part... d'autre part..., plus... plus...* projettent une seconde occurrence symétrique de la première. Les structures clivées *c'est... qui...* et *c'est... que...* focalisent un constituant et contraignent la suite de l'énoncé à une proposition relative. L'amorce *Ce que j'apprécie chez elle, c'est \_\_* appelle un nom abstrait désignant une qualité : *sa franchise, sa rigueur, son humour*.

Le cinquième ressort est celui de la concordance des temps dans le discours rapporté et la subordonnée complétive. La matrice *Il m'a dit qu'il \_\_* projette, à la voix rapportée, un imparfait, un conditionnel ou un plus-que-parfait selon la perspective aspectuelle visée : *qu'il viendrait, qu'il était parti, qu'il travaillait à ce projet*. La cohérence temporelle constitue une contrainte prédictive forte, dont la violation est

immédiatement perçue comme fautive.

**Exemple 1 — Hypothèse et concordance :** Dans *Si j'avais su, \_\_*, le plus-que-parfait dans la protase projette une apodose au conditionnel passé : *je ne serais pas venu, j'aurais agi autrement, je t'aurais prévenu*. Ce schéma est l'un des plus contraignants de la syntaxe française.

**Exemple 2 — Verbe à double complémentation :** L'énoncé *L'enseignante a demandé aux élèves de \_\_* projette un infinitif : *rendre leurs devoirs, se taire, réfléchir avant de répondre*. Le verbe *demandé* à *quelqu'un* de sélectionne strictement une proposition infinitive.

**Exemple 3 — Relative restrictive :** Dans *Le dossier que \_\_*, le pronom relatif *que* occupe une fonction d'objet direct et projette un sujet suivi d'un verbe transitif : *que le directeur a validé, que j'ai consulté hier, que tu m'as remis*. L'accord du participe passé avec l'antécédent en position d'objet direct antéposé constitue une contrainte formelle prédictive.

**Exemple 4 — Concessive et retournement :** L'énoncé *Même si la route était verglacée, \_\_* projette une principale qui renverse l'attente négative : *nous sommes arrivés à l'heure, le convoi a poursuivi sa progression, personne n'a renoncé au voyage*.

**Exemple 5 — Gradation temporelle :** La séquence *Dans un premier temps, \_\_* appelle un complément posant l'étape initiale d'un raisonnement, et projette un *dans un second temps* ou *par la suite* qui en prolonge l'architecture argumentative. Les connecteurs temporels fonctionnent comme des balises qui structurent l'avancée du discours.

**Exemple 6 — Clivée focalisante :** L'amorce *C'est précisément là que \_\_* projette un verbe introduisant une prise de position ou un constat : *le problème se pose, les choses se compliquent, l'enjeu devient crucial*. La structure clivée oriente vers une suite à valeur argumentative forte.

**Exemple 7 — Négation et restriction :** L'énoncé *Je ne veux pas que \_\_* appelle obligatoirement une complétive au subjonctif : *tu partes maintenant, cette discussion s'envenime, nous perdions notre temps*. L'amorce *Il ne reste plus qu'à \_\_* projette à l'inverse un infinitif : *attendre les résultats, prendre une décision, accepter la situation*.

**Exemple 8 — Corrélation d'intensité :** L'amorce *Plus il travaillait, \_\_* projette une seconde proposition introduite par *plus* ou *moins* qui établit une relation de covariation : *plus il progressait, moins il doutait, plus ses résultats s'amélioraient*. La corrélation impose un

parallélisme strict entre les deux volets.

## La prédiction pragmatique en français

La prédiction pragmatique concerne les unités larges de l'organisation discursive : le scénario d'interaction, la séquence conversationnelle, le rituel social, le schéma narratif. À cette échelle, l'anticipation ne porte plus sur un mot isolé ni sur une structure phrastique, mais sur le type d'acte de langage attendu, sur la forme qu'appellera la réponse, et sur la position que chaque tour occupe dans un script culturellement partagé. Le français, avec sa codification fine des registres et sa tradition rhétorique, offre un terrain d'observation particulièrement riche.

Le premier axe concerne les rituels d'ouverture et de clôture des échanges. Entrer dans un commerce sans saluer le personnel constitue, en France, un manquement perçu comme notable. L'amorce *Bonjour madame* appelle une réponse symétrique *Bonjour monsieur* avant toute autre transaction. Le passage au registre familier est lui-même codifié : la proposition *On peut se tutoyer, non ?* reconfigure l'espace des attentes pour la suite de l'échange. La clôture obéit à des formules stabilisées : *Je vous en prie, À bientôt, Bonne journée, Passez une bonne fin de journée*, dont l'absence serait ressentie comme abrupte.

Le deuxième axe est celui du repas partagé et de l'hospitalité. Dans la culture française, l'invitation à table active un scénario d'une forte stabilité. L'hôte ouvre par *À table !* ou *Passons à table*, qui appelle l'installation des convives. Le service d'un plat est fréquemment accompagné du souhait *Bon appétit !*, auquel la réponse canonique est *Merci, à vous aussi* ou *Merci, de même*. L'offre de resservir s'exprime par *Je te ressers ?* ou *Encore un peu ?*, et appelle typiquement un refus courtois suivi d'une acceptation partielle : *Non merci, j'ai déjà très bien mangé*, puis *Allez, un petit peu alors, c'est si bon*. Le compliment adressé à la cuisine (*C'est délicieux, C'est un régal*) est attendu et structurant de l'interaction.

Le troisième axe porte sur les félicitations, condoléances et autres formules de circonstance. À l'annonce d'une heureuse nouvelle, *Toutes mes félicitations !* ou *Je suis très heureux pour toi* appellent *Merci beaucoup, cela me touche* ou *Merci, c'est gentil à toi*. En contexte de deuil, *Je vous présente mes sincères condoléances* ou *Toute ma sympathie dans cette épreuve* appellent *Je vous remercie, votre soutien m'est précieux*. Les formules d'anniversaire, de réussite ou de rétablissement (*Bon rétablissement, Prompt rétablissement, Mes meilleurs vœux*) sont hautement

codifiées et admettent très peu de variation.

Le quatrième axe concerne les stratégies d'atténuation et la politesse indirecte. Une demande directe — *Passe-moi le pain* — est acceptable en contexte familial mais serait perçue comme abrupte en contexte formel. La forme attendue est : *Est-ce que tu pourrais me passer le pain, s'il te plaît ?* ou, en registre soutenu : *Pourriez-vous me passer le pain, je vous prie ?*. Le conditionnel de politesse (*pourriez-vous, auriez-vous, voudriez-vous*), les tournures introductrices (*je me permets, si je puis me permettre, à toutes fins utiles*) et les formules de clôture (*s'il vous plaît, je vous en serais reconnaissant*) constituent un arsenal prédictif stable de la requête polie.

Le cinquième axe concerne les scénarios narratifs et les tournures d'introduction. *Figurez-vous que* — projette une anecdote surprenante ou significative. *Pour faire simple*, — annonce un résumé condensé. *À ce propos*, — signale un glissement thématique. *Pour en revenir à ce que je disais*, — marque une reprise après digression. Ces marqueurs opèrent comme des balises discursives qui structurent l'attente de l'interlocuteur et orientent son interprétation.

Le sixième axe regroupe les particules et interjections qui modulent le tour de parole. *Bon...* en ouverture de tour signale souvent une transition ou une conclusion ; *Enfin...* introduit une concession ou une nuance ; *Voilà* clôt ou ponctue un énoncé ; *Quand même* marque l'insistance ou la désapprobation feutrée. Ces petits mots, souvent négligés dans les grammaires scolaires, constituent des leviers prédictifs puissants pour anticiper le déroulement de la conversation.

**Exemple 1 — Rencontre formelle :** *Enchanté de faire votre connaissance* projette *Tout le plaisir est pour moi*, suivi d'un échange de présentations professionnelles. Le vouvoiement contraint l'ensemble du registre pour la durée de l'échange et n'est relâché qu'à l'issue d'une négociation explicite.

**Exemple 2 — Invitation informelle :** *On se fait un café ? / On prend un verre ?* projette soit une acceptation (*Avec plaisir !, Volontiers !*), soit un refus motivé (*C'est gentil, mais là je suis vraiment débordé, on remet ça à la semaine prochaine ?*). Le refus sans motivation serait culturellement marqué et pourrait être interprété comme une mise à distance.

**Exemple 3 — Réception d'un compliment :** *Tu as une très belle robe !* appelle typiquement une réponse minorative : *Oh, c'est une vieille robe, je l'ai depuis des années*, ou une redirection : *C'est gentil, je l'ai trouvée*

*en soldes*. La réponse *Oui, je sais, elle me va bien* serait pragmatiquement atypique dans la norme française, qui valorise la modestie conversationnelle.

**Exemple 4 — Excuse et minimisation :** *Je suis vraiment désolé pour ce retard* projette *Ce n'est rien, Ne vous en faites pas*, ou *Il n'y a pas de mal*. La minimisation par l'interlocuteur constitue une convention stable qui referme la séquence d'excuse.

**Exemple 5 — Clôture téléphonique :** *Bon, je vais te laisser* projette *D'accord, à bientôt !*, suivi éventuellement de *Passe une bonne soirée, Bises, Je t'embrasse, À très vite*. Raccrocher sans passer par cette séquence serait perçu comme brutal, même entre proches.

**Exemple 6 — Demande de renseignement en situation :** *Excusez-moi de vous déranger, auriez-vous l'heure, s'il vous plaît ?* projette une réponse courtoise : *Il est midi et demi*, ou *Je suis désolé, je n'ai pas de montre*. La mise en scène explicite de la gêne (*excusez-moi de vous déranger*) précède presque systématiquement la demande adressée à un inconnu.

**Exemple 7 — Requête professionnelle atténuée :** *Je me permets de vous solliciter au sujet de* — projette l'exposition d'un motif, suivie d'une demande formulée au conditionnel : *serait-il possible de..., pourriez-vous m'indiquer..., j'aimerais savoir si....* Les épistolaires professionnels se referment par une formule attendue : *Je vous remercie par avance de votre attention* et *Je vous prie d'agréer, Madame, Monsieur, mes salutations distinguées*.

**Exemple 8 — Débat d'opinion :** L'amorce *Je ne suis pas tout à fait d'accord avec toi sur ce point* projette une objection développée, introduite souvent par *parce que* ou *dans la mesure où*, et close par une réouverture : *mais je peux comprendre ton point de vue, mais c'est un débat légitime*. Le désaccord frontal non atténué serait perçu comme agressif dans la plupart des registres conversationnels.

Ces trois niveaux de prédiction — sémantique, syntaxique et pragmatique — ne constituent pas des strates autonomes. Ils interagissent constamment : le choix d'un connecteur suppose une anticipation lexicale, l'ouverture d'un champ sémantique engage une structure phrastique attendue, et la reconnaissance d'un rituel social oriente l'ensemble des choix morphosyntaxiques et lexicaux qui s'enchaîneront. C'est dans cette circulation continue que se loge la spécificité prédictive du français.

## Rhétorique et prédiction en français

Aux trois niveaux d'analyse qui viennent d'être décrits (sémantique, syntaxique et pragmatique) s'ajoute une dimension dont l'importance, pour le français, ne saurait être sous-estimée : la dimension rhétorique. Par rhétorique, on entend ici l'ensemble des dispositifs formels et des architectures argumentatives que la tradition discursive française a codifiés avec une rigueur particulière, depuis l'éloquence classique du Grand Siècle jusqu'aux exercices contemporains de la dissertation, du commentaire composé, de l'éditorial et de la note de synthèse. Ces dispositifs ne relèvent ni strictement du lexique, ni strictement de la syntaxe, ni strictement de la pragmatique : ils les mobilisent conjointement pour construire des macro-structures discursives dont la prévisibilité est, dans l'espace culturel francophone, particulièrement stabilisée.

La thèse que l'on voudrait avancer ici est simple : la rhétorique française constitue une ressource prédictive à part entière, dont la maîtrise conditionne en profondeur la capacité d'un locuteur (humain ou artificiel) à produire et à comprendre un discours jugé cohérent, élégant et pertinent. Ignorer cette dimension reviendrait à réduire la grammaire prédictive du français à une mécanique phrastique incapable de rendre compte de la manière dont les énoncés s'organisent en séquences argumentatives, en développements démonstratifs ou en récits structurés.

Le premier axe de prédiction rhétorique concerne les architectures argumentatives canoniques. Le français a hérité, notamment par la voie scolaire et universitaire, d'une tripartition argumentative d'une stabilité remarquable : l'énoncé d'une thèse, la formulation d'une antithèse et la proposition d'une synthèse. Cette armature, souvent intériorisée dès le lycée à travers l'exercice de la dissertation, structure non seulement les productions écrites académiques, mais également les éditoriaux journalistiques, les interventions politiques, les plaidoiries judiciaires et même certaines conversations argumentées du quotidien. La réfutation et la concession sont deux cas de figures systématiquement présents dans les discours argumentatifs. Tout lecteur ou interlocuteur averti peut en effet prévoir à la fois l'apparition de séquences réfutatives ou concessives dans tout discours argumentatif. Ainsi l'apparition d'une amorce telle que *On objectera sans doute que* \_\_\_ projette presque inévitablement une antithèse qui sera elle-même dépassée par un *Il faut néanmoins*

*reconnaître que* \_\_\_ ou un *C'est pourquoi il convient plutôt de considérer que* \_\_\_. L'auditeur ou le lecteur francophone averti anticipe ces mouvements avant même qu'ils ne se réalisent, et l'économie argumentative du texte en tire une grande partie de sa lisibilité. De la même manière, l'apparition dans le discours de « certes ... » qui annonce une concession à la thèse opposée, ne laisse aucun doute sur la suite qui ne peut être qu'une remise en question partielle ou radicale de ce qui a été concédé au départ. Le « Mais... » dans le système concessif annonce inévitablement l'argument qui, d'un côté appuie la thèse, et d'un autre côté, rend négligeable ce qui a été concédé. Bref, le couple « certes...mais... » sont à la fois annonciateur d'un ordre argumentatif et hiérarchise la validité des arguments : ceux qui appuient la thèse et sur lesquels on met l'emphase, et ceux concédés et qui sont en même temps relégués au second plan.

Le deuxième axe concerne les figures d'insistance et les figures de construction. La rhétorique française privilégie certaines figures dont la récurrence crée des attentes prédictives précises. L'anaphore, qui consiste à reprendre un même terme en tête de propositions successives, projette sa propre continuation : l'amorce *La République, c'est la justice ; la République, c'est* \_\_\_ appelle *l'égalité, la fraternité*, ou un autre terme du champ républicain. Le chiasme (*Il faut manger pour vivre et non vivre pour manger*) crée une attente de symétrie inversée que l'interlocuteur complète mentalement avant même son énonciation complète. Le parallélisme (*Plus on donne, plus on reçoit*) impose une structure miroir dont la rupture serait immédiatement perçue. Le tricolon, enfin, cette figure ternaire chère à la tradition française (*veni, vidi, vici ; liberté, égalité, fraternité ; le travail, la famille, la patrie*), installe une attente de trois éléments dont l'absence du troisième produit un effet de suspension immédiatement ressenti.

Le troisième axe porte sur les connecteurs argumentatifs à valeur rhétorique. Le français dispose d'un appareil de transitions d'une richesse exceptionnelle, dont chaque élément porte une instruction procédurale précise sur la suite du raisonnement. *Non seulement... mais encore...* annonce une gradation à renforcement. *D'abord... ensuite... enfin...* balise une énumération ordonnée. *D'une part... d'autre part...* distribue le raisonnement en deux volets équilibrés. *Or* introduit un argument décisif qui fait basculer le raisonnement vers sa conclusion. *En effet* justifie rétrospectivement ce qui vient d'être avancé. *Ainsi et de sorte que* tirent une conséquence. *En définitive, au demeurant, au bout du compte, tout bien considéré* signalent l'approche de la conclusion.

Ces connecteurs, souvent rangés sous la bannière un peu réductrice des « mots de liaison » dans l'enseignement, fonctionnent en réalité comme des opérateurs prédictifs de haut niveau qui orientent l'interprétation du discours entier.

Le quatrième axe concerne les scénarios discursifs propres aux genres textuels. Chaque genre — dissertation, commentaire, éditorial, essai, conférence, préface, nécrologie, éloge funèbre, discours de remise de prix — suit en français une architecture dont la prévisibilité est tellement forte qu'elle relève presque du rituel. La dissertation classique s'ouvre typiquement sur une mise en contexte (*De tout temps, les hommes se sont interrogés sur* \_\_\_), se poursuit par l'énoncé d'une problématique (*On peut dès lors se demander si* \_\_\_), se prolonge par l'annonce d'un plan (*Dans un premier temps, nous examinerons* \_\_\_ ; *puis nous verrons que* \_\_\_ ; *avant de conclure sur* \_\_\_) et se referme sur une ouverture (*Il reste cependant à interroger* \_\_\_). L'éloge funèbre suit un autre scénario : évocation biographique, énumération des mérites, expression du manque, consolation des proches. L'éditorial journalistique, enfin, déploie une actualité, en dégage un enjeu, adopte une position et la justifie par des arguments. Ces armatures textuelles sont si prévisibles que leur violation produit un effet stylistique marqué, parfois recherché à des fins de rupture.

Le cinquième axe, enfin, concerne la rhétorique de la politesse écrite et des formules de cadrage institutionnel. Le français entretient une tradition épistolaire d'une codification exceptionnelle. L'ouverture d'une lettre officielle par *Madame, Monsieur* projette une architecture entière : un paragraphe de cadrage introduit par *J'ai l'honneur de* \_\_\_ ou *Je me permets de* \_\_\_, un développement exposant l'objet de la demande, une clause de conclusion introduite par *Dans l'attente de votre réponse* \_\_\_ et une formule finale stabilisée : *Je vous prie d'agréer, Madame, Monsieur, l'expression de mes salutations distinguées* ou, dans un registre plus élevé, *Je vous prie de croire, Madame, Monsieur, en l'assurance de ma considération la plus distinguée*. Chacun de ces éléments appelle les autres avec une quasi-nécessité, au point que leur absence ou leur interversion est immédiatement perçue comme une faute de convenance, indépendamment de toute question grammaticale.

**Exemple 1 — Concession rhétorique :** L'amorce *Il serait tentant de croire que* \_\_\_ projette une proposition que le locuteur entend précisément

contester, et appelle son retournement par un *Pourtant*, un *Or*, ou un *C'est oublier que* \_\_\_. L'auditeur francophone anticipe le geste de réfutation avant sa formulation explicite.

**Exemple 2 — Gradation ternaire :** Dans un discours politique, l'amorce *Nous défendrons cette cause avec courage, avec détermination,* \_\_\_ projette un troisième terme à valeur climatique : *et avec la plus ferme conviction, et jusqu'au bout, et quoi qu'il en coûte*. La clôture binaire serait ressentie comme inachevée.

**Exemple 3 — Question rhétorique :** L'énoncé *Qui oserait prétendre aujourd'hui que* \_\_\_ projette une proposition présentée comme si manifestement indéfendable qu'elle n'appelle aucune réponse effective. Le locuteur enchaînera par *Personne, évidemment* ou par un silence entendu suivi d'un développement qui tire argument de cette évidence supposée.

**Exemple 4 — Structure d'éditorial :** L'ouverture *Depuis plusieurs semaines, le débat public s'enflamme autour de* \_\_\_ projette une architecture canonique : exposition des positions en présence, prise de distance critique, énoncé d'une position propre et conclusion ouvrant sur un enjeu plus large. Le lecteur averti anticipe chacun de ces mouvements avant qu'ils ne se déploient.

**Exemple 5 — Clôture académique :** La formule *Au terme de cette analyse, il apparaît que* \_\_\_ projette une reformulation synthétique des acquis de la démonstration, suivie d'une ouverture prospective introduite par *Il resterait néanmoins à examiner* \_\_\_ ou *Une telle perspective invite à prolonger la réflexion vers* \_\_\_. La conclusion sèche, sans ouverture finale, serait perçue comme tronquée dans la tradition académique française.

**Exemple 6 — Éloquence du lieu commun :** Dans un registre oratoire, l'amorce *Comme chacun sait* \_\_\_ ou *Nul n'ignore que* \_\_\_ projette l'énoncé d'une proposition que le locuteur présente comme consensuelle afin de l'utiliser comme prémisse d'un raisonnement ultérieur. L'interlocuteur francophone perçoit immédiatement cette stratégie de présupposition et anticipe la conclusion qui en sera tirée.

Ces dispositifs rhétoriques opèrent à une échelle supérieure à celle de la phrase, mais ils entretiennent avec les trois niveaux précédents une relation d'interpénétration constante. La figure d'anaphore, par exemple, mobilise simultanément une contrainte sémantique (cohésion lexicale), une contrainte

syntaxique (parallélisme des structures) et une contrainte pragmatique (effet d'insistance orienté vers un auditoire). De même, le connecteur *en définitive* active à la fois une attente lexicale (vocabulaire conclusif), une attente syntaxique (proposition bilan) et une attente rhétorique (clôture du raisonnement). La grammaire prédictive du français ne peut donc se contenter d'aligner les niveaux : elle doit rendre compte de leur convergence dans des architectures discursives dont la tradition rhétorique française offre les modèles les plus stabilisés.

C'est précisément cette dimension rhétorique, souvent invisible aux grammaires phrastiques comme aux modèles statistiques standards, qui confère au discours français une part notable de sa prévisibilité et, par conséquent, une part essentielle de sa difficulté pour les apprenants comme pour les systèmes automatiques de traitement du langage.

## Grammaire prédictive et modèles de langage

Les analyses qui précèdent débordent le cadre de la seule description grammaticale ou de la didactique des langues. La grammaire prédictive, telle qu'elle a été esquissée ici pour le français, livre un cadre heuristique dont les retombées intéressent directement la conception des systèmes d'intelligence artificielle dédiés au traitement du langage naturel.

Il est notable que les architectures neuronales qui sous-tendent les grands modèles de langage contemporains reposent, dans leur principe même, sur un mécanisme prédictif. Qu'il s'agisse de modèles autorégressifs, de modèles par diffusion ou de systèmes hybrides récents, leur fonctionnement consiste à estimer, à chaque pas de génération, la distribution de probabilité de l'unité textuelle suivante conditionnée à un contexte. La convergence entre cette mécanique computationnelle et la prédiction linguistique telle que la décrivent les sciences cognitives n'est pas anecdotique : elle signale une parenté structurelle entre l'anticipation humaine et les algorithmes d'apprentissage automatique.

Cependant, les grands modèles actuels ont été majoritairement entraînés sur des corpus à dominance anglophone et traitent en général les autres langues comme des variantes périphériques d'une matrice commune. Cette homogénéisation produit des systèmes formellement opérants mais souvent culturellement approximatifs, capables de générer des énoncés grammaticalement acceptables sans pour autant respecter les régimes de politesse,

les scénarios d'interaction ou les conventions rhétoriques propres à la langue cible. Un modèle qui répondrait à un courriel professionnel français par un simple *Salut !* suivi d'un texte non structuré produirait un énoncé compréhensible mais socialement inadéquat.

La grammaire prédictive du français, telle qu'elle vient d'être formulée, offre à cet égard trois apports concrets à la conception de modèles de langage plus finement adaptés.

Le premier concerne la représentation des unités et la tokenisation. Une segmentation non informée, héritée de l'anglais, tend à fragmenter des ensembles morphologiquement solidaires en français (préfixes verbaux, terminaisons accordées, liaisons obligatoires) et à diluer l'information grammaticale dans un bruit statistique. Une segmentation guidée par la description prédictive préserverait la cohésion des familles lexicales, la stabilité des collocations et l'intégrité des locutions figées comme autant de niveaux d'organisation à respecter.

Le deuxième apport porte sur le traitement des marqueurs fonctionnels — connecteurs, pronoms clitiques, modes verbaux, chaînes d'accord. Ces éléments opèrent en français comme des opérateurs prédictifs à longue portée : ils conditionnent, sur plusieurs mots ou plusieurs propositions, l'ensemble des continuations probables. Un modèle linguistiquement informé les reconnaîtrait comme des points de bifurcation structurant l'espace de décision, plutôt que de compter sur la seule redécouverte statistique de leurs régularités.

Le troisième apport, peut-être le plus décisif, relève de la dimension pragmatique et culturelle. Un modèle qui prolongerait un *Bonjour monsieur, je me permets de vous écrire au sujet de votre dossier* par un registre familier ou par un style abrupt manquerait gravement aux attentes interactionnelles propres au français écrit soutenu. De même, un modèle qui ignorerait la séquence *offre — refus poli — insistance — acceptation modérée* dans un scénario d'hospitalité produirait une réponse linguistiquement correcte mais socialement fautive. Ces décalages, souvent invisibles aux métriques quantitatives d'évaluation, pèsent lourdement sur l'acceptabilité réelle des productions en contexte francophone.

L'exigence d'adaptation culturelle ne se résume pas à une localisation de surface : traduction d'interfaces, ajustement des unités de mesure ou des conventions typographiques. Elle implique de reconnaître que chaque langue encode, dans la texture même de ses structures, une représentation du monde, une

hiérarchie des rapports sociaux et un régime d'interaction. L'importance du *bonjour* en ouverture, la pratique codifiée du compliment gastronomique, la prévalence des tournures atténuées dans la requête, la densité des formules épistolaires, la vitalité du patrimoine proverbial et des références littéraires : autant d'éléments qui ne constituent pas un vernis pragmatique apposé à une grammaire neutre, mais la fibre même de la prévisibilité du français.

Un modèle culturellement adapté est, au fond, un modèle qui a appris à anticiper non seulement ce qui est dit, mais aussi ce qui, dans un contexte donné, doit être insinué, atténué ou signalé par la seule architecture formelle de l'énoncé. C'est à ce prix que l'intelligence artificielle accédera à une compétence linguistique véritablement située — c'est-à-dire apte non seulement à produire des phrases conformes, mais à s'inscrire dans les mille régularités implicites qui font qu'une langue est une communauté et non un simple code.

## Conclusion

Cet article a cherché à poser les jalons d'une grammaire prédictive du français, en articulant ses fondements théoriques, ses principales opérations et ses extensions possibles. La démarche reposait sur un déplacement central : passer d'une grammaire de la règle à une grammaire de la probabilité, où la compétence linguistique se mesure moins à la connaissance de paradigmes explicites qu'à la capacité d'anticiper, à chaque pas de l'énoncé, les continuations les plus plausibles.

Trois niveaux d'analyse ont été distingués et documentés. Au plan sémantique, la productivité des affixes dérivationnels, la cohésion des champs lexicaux et la cristallisation des locutions figées constituent autant d'ancrages qui orientent l'anticipation du mot suivant. Au plan syntaxique, la valence verbale, les pronoms clitiques antéposés, les connecteurs logiques, les constructions corrélatives et la concordance des temps déploient des contraintes à longue portée qui projettent la structure phrastique à venir. Au plan pragmatique, les rituels de civilité, les scripts d'hospitalité, les stratégies d'atténuation et les marqueurs discursifs organisent la prévisibilité des scénarios interactionnels et narratifs. Ces trois plans, on l'a souligné, ne fonctionnent pas isolément : ils s'entrelacent en une dynamique intégrée qui fait de la prédiction le principe organisateur de l'activité langagière.

Sur un plan didactique, cette approche invite à repenser l'enseignement du français : moins comme la transmission d'un corps de règles à mémoriser que comme la construction progressive d'une compétence anticipatoire, sensible aux régularités probabilistes qui régissent l'usage en contexte. Apprendre le français, dans cette optique, c'est apprendre à prévoir — à prévoir le mot qui va venir, la structure qui va se déployer, la réaction qui va répondre.

Sur un plan plus prospectif, l'articulation entre linguistique prédictive et intelligence artificielle dessine l'horizon d'une génération nouvelle de modèles de langage que l'on pourrait qualifier de *linguistiquement pertinents*. Pertinents d'abord, parce qu'ils reconnaîtraient la singularité typologique des langues qu'ils traitent, au lieu de les réduire à une matrice anglophone implicite. Pertinents ensuite, parce qu'ils combinerait l'apprentissage distributionnel massif avec des connaissances linguistiques explicites — affixes productifs, connecteurs modaux, chaînes d'accord, scripts pragmatiques, mobilisables comme biais inductifs ou modules de contrôle. Pertinents enfin, parce qu'ils s'évalueraient selon des critères non plus seulement formels, mais culturellement situés : adéquation pragmatique, maîtrise des registres, respect des conventions interactionnelles.

Le français, par la dissymétrie de son oral et de son écrit, la richesse de son système déterminatif, la finesse de ses codes pragmatiques et la densité de son héritage rhétorique, constitue à cet égard un cas d'étude exemplaire. Ce que la linguistique prédictive parvient à formaliser de cette langue peut, *mutatis mutandis*, être transposé à d'autres systèmes linguistiques dont la spécificité typologique et culturelle demeure aujourd'hui insuffisamment captée par les modèles dominants.

La grammaire prédictive n'est donc pas un point d'arrivée, mais un seuil. Elle ouvre un programme de recherche dans lequel didactique des langues et conception des intelligences artificielles cessent d'être des champs parallèles pour se reconnaître comme deux faces d'un même projet : comprendre, formaliser et transmettre ce qui, dans chaque langue, rend le sens possible et l'interaction humaine intelligible.

## Références

- Amos, R. M., Seeber, K. G., & Pickering, M. J. (2022). Prediction during simultaneous interpreting: Evidence from the visual-world paradigm. *Cognition*, 220, 104987.
- Babineau, M., Havron, N., Dautriche, I., de Carvalho, A., & Christophe, A. (2023). Learning to predict and predicting to learn: Before and beyond the syntactic bootstrapper. *Language acquisition*, 30(3-4), 337-360.
- Barbazan, M. (2008). Principes d'une grammaire prédictive du discours (français langue étrangère et maternelle). In *Congrès Mondial de Linguistique Française* (p. 053). EDP Sciences.
- Berthele, R., Lenz, P., & Peyer, E. (2022). Predicting foreign language skills based on first languages: The role of lexical distance and relative morphological complexity. *Poznan Studies in Contemporary Linguistics*, 58(3), 419-448.
- Bigi, B. (2023, April). An analysis of produced versus predicted French Cued Speech keys. In *10th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics* (pp. 24-28).
- Blais, G. (2023). Le traitement de l'accord de l'adjectif et du déterminant en modalité auditive: une étude de potentiels évoqués chez les adolescents québécois. Mémoire de Maîtrise. Faculté des arts et des sciences – Département de linguistique et de traduction. Université de Montréal.
- Botes, E., Dewaele, J. M., Greiff, S., & Goetz, T. (2024). Can personality predict foreign language classroom emotions? The devil's in the detail. *Studies in Second Language Acquisition*, 46(1), 51-74.
- Cattani, M. T. (2026). *L'influence de la connaissance du vocabulaire et de la grammaire sur la compréhension écrite en français LE* (Master's thesis).
- Da Cunha, Y., & Abeillé, A. (2022). L'alternance actif/passif en français parlé: un modèle statistique. *Langue française*, 216(4), 63-80.
- Dalola, A. (2022). Introduction to Special Issue on French Variation in Digital Media. *Journal of French Language Studies*, 32(2), 115-119.
- De Fino, V., Fontan, L., Pinquier, J., Barcat, C., Ferrané, I., & Detey, S. (2022, June). Mesures automatiques de parole non-native: exploration pilote d'un corpus d'apprenants japonais de français et différenciation de niveaux. In *34èmes Journées d'Études sur la Parole (JEP 2022)* (pp. 1-10).
- Desrochers, A., Paivio, A., & Desrochers, S. (1989). L'effet de la fréquence d'usage des noms inanimés et de la valeur prédictive de leur terminaison sur l'identification du genre grammatical. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 43(1), 62.
- Destruel, E. (2023). Processing pragmatic inferences in L2 French speakers. *Second language research*, 39(4), 969-995.
- Dündar, O. İ. (2024). Utilisation Potentielle De Chatgpt Dans L'apprentissage Des Langues Etrangères: Exploration Des Possibilités Selon Les Niveaux Langagiers Du CECRL. *Kahramanmaraş Sütçü İmam Üniversitesi Sosyal Bilimler Dergisi*, 21(1), 63-75.
- E Bosch, J., & Foppolo, F. (2022). Predictive processing of grammatical gender in bilingual children: The effect of cross-linguistic incongruency and language dominance. *Lingue e linguaggio*, (1), 5-27.
- Ferrand, L., & New, B. (2023). Prédiction lexicale et traitement du français: apports des approches probabilistes. *Langue française*, 218(2), 45-68.
- Gavard, E. (2024). *Le rôle de la prédiction sémantique et syntaxique et de l'apprentissage statistique dans la lecture normale et dans la dyslexie développementale* (Doctoral dissertation, Aix-Marseille Université).
- Gavard, E., & Ziegler, J. C. (2023). The effects of semantic and syntactic prediction on reading aloud. *Experimental Psychology*.
- Grisot, C., Blochowiak, J., & Puskás, G. (2022). Predicting verbal mood in complement clauses: the role of the semantic category, the tense and the grammatical aspect of the matrix verb. *Glossa: a journal of general linguistics*, 7(1).
- Grisot, C., Blochowiak, J., & Puskás, G. (2022). Predicting verbal mood in complement clauses: the role of the semantic category, the tense and the grammatical aspect of the matrix verb. *Glossa: a journal of general linguistics*, 7(1).
- Guidère, M. (2015). *La Linguistique prédictive : de la cognition à l'action*. Paris: Éditions L'Harmattan. Collection Traductologie.
- Guidère, M. (2025). La linguistique prédictive : concepts, méthodes et applications. *Journal of Applied Research in Human & Social Sciences*, 1(1), 72-89.
- Guillot, N. (2022, November). La notion de Modèle en Grammaire Générative. In *Séminaire du Laboratoire*

MoDyCo" *Epistémologie des Sciences du Langage*".

Ito, A., Bautista, A., & Martin, C. (2025). Cognate facilitation effect on verb-based semantic prediction in L2 is modulated by L2 proficiency. *Bilingualism: Language and Cognition*, 28(3), 684-696.

Klasen, L., Ugen, S., Dording, C., Fayol, M., & Weth, C. (2024). Do learners need semantics to spell syntactic markers? Plural spellings in real vs. pseudowords in a French L2 setting. *Reading and Writing*, 37(1), 129-145.

Koch, E. M., Bulté, B., Housen, A., & Godfroid, A. (2023). The predictive processing of number information in subregular verb morphology in a first and second language. *Applied Psycholinguistics*, 44(5), 750-783.

Kronning, H. (2023). Modalité, typologie et cognition: la construction devoir+ INFINITIF comme périphrase future et évidentielle. *Acta Linguistica Hafniensia*, 55(sup1), 24-43.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1), 32-59.

Larrivée, P., & Pinzin, F. (2025). L'évolution de l'expression du sujet impersonnel dans l'histoire du français. *Bulletin de la Société de Linguistique de Paris*.

Le Diagon, S. (2024). *Quels sont les mécanismes impliqués dans la relation entre compétences académiques émergentes et ultérieures? Une étude longitudinale de la maternelle au CM2* (Doctoral dissertation, Université Claude Bernard-Lyon I).

Le Normand, M. T., & Thai-Van, H. (2022). The role of function words to build syntactic knowledge in French-speaking children. *Scientific reports*, 12(1), 544.

Lee, A. H., & Lyster, R. (2023). The effects of different types of form-focused instruction on the acquisition of grammatical gender by second language learners of French. *Language awareness*, 32(3), 508-529.

Lin, Q. (2022). Anxiety and self-efficacy in Chinese international students' L3 French learning with L2 English and L3 French. *Frontiers in Psychology*, 13, 998536.

Loucif, S. (2023). *Les phénomènes de transfert en acquisition du genre grammatical en français langue seconde* (Doctoral dissertation, Université Paris Cité).

MacKay, E., Chen, X., & Deacon, S. H. (2023). Longitudinal predictors of French word reading

difficulties among French Immersion children. *Annals of dyslexia*, 73(1), 73-89.

Mihaylov, V., & Shtedritski, A. (2024, November). What an elegant bridge: Multilingual LLMs are biased similarly in different languages. In *Proceedings of the Fourth Workshop on Multilingual Representation Learning (MRL 2024)* (pp. 22-29).

Moeschler, J., & Reboul, A. (2019). *La pragmatique aujourd'hui : une nouvelle science de la communication*. Seuil.

North, K., Zampieri, M., & Shardlow, M. (2023). Lexical complexity prediction: An overview. *ACM Computing Surveys*, 55(9), 1-42.

Palasis, K., Faure, R., & Meunier, F. (2023). Wh-in-situ in child French: Deictic triggers at the syntax-semantics interface. *Journal of French Language Studies*, 33(3), 273-298.

Pineau, L., Montermini, F., & Thuilier, J. (2024, August). The identification and processing of neologisms. An experimental approach to neological intuition in French. In *21st International Morphology Meeting (IMM21)*.

Schlechter, J. (2023). Prediction in bilingual sentence processing: How prediction differs in a later learned language from a first language. *Bilingualism: Language and Cognition*, 26(2), 253-267.

Schlechter, J. (2023). Prediction in bilingual sentence processing: How prediction differs in a later learned language from a first language. *Bilingualism: Language and Cognition*, 26(2), 253-267.

Thebaud, M., Antoine, J. Y., Allegre, W., Tsimba, V., Bossard, I., Crochetet, M., ... & Pouplin, S. (2025). The Predict4All open-source word prediction and word correction engine in French language: a feasibility study. *Disability and Rehabilitation: Assistive Technology*, 20(5), 1557-1565.

Torres-Martínez, S. (2023). Grammaire agentielle cognitive de constructions. Explorations sémiotico-linguistiques des origines de la représentation incarnée. *Signata. Annales des sémiotiques/Annals of Semiotics*, (14).

Torterat, F., Gomila, C., Bazile, S., Cwiczynski, M., Gennaï, A., & Soulié, A. (2024). Les glossaires comme supports d'élaboration grammaticale en contexte de formation. *Pratiques. Linguistique, littérature, didactique*, (201-202).

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P.,

Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3), 443–467.

Vianne, L., Dupont, Y., & Barré, J. (2023, December). Gender Bias in French Literature. In *Conference on Computational Humanities Research CHR2023*.

Zribi-Hertz, A. (2022). Accord, genre et prédiction : ressources grammaticales du français dans le traitement en ligne. *Langue française*, 216(4), 89–112.