

LA PRÉDICTION LINGUISTIQUE EN CHINOIS : ASPECTS THÉORIQUES ET PRATIQUES

Linguistic Prediction in Chinese: Theoretical and Practical Aspects

Min SUN. Professeure à l'Université du Québec à Montréal (UQÀM, Canada)
GRÉLIA (Groupe de Recherche et d'Étude en Langues et Intelligence Artificielle)
sun.min@uqam.ca

Mathieu GUIDERE. Professeur à l'Université de Paris. Directeur de recherches (France)
INSERM (Institut National de la Santé et de la Recherche Médicale)
mathieu.guidere@u-paris.fr

Résumé : Cet article prolonge les fondements de la linguistique prédictive (Guidère, 2015, 2025) en les appliquant au chinois mandarin. Il montre que la langue chinoise, caractérisée par une morphologie isolante, un ordre des mots strict, une forte dépendance au contexte et un tissu culturel dense, constitue un terrain privilégié pour l'analyse prédictive. Après un rappel du cadre théorique issu des grands modèles de langage, l'article développe une *grammaire prédictive du chinois* articulée autour de trois niveaux : la *prédiction sémantique* (anticipation du mot suivant à partir des cooccurrences lexicales et des radicaux sémantiques), la *prédiction syntaxique* (anticipation du syntagme et de la phrase suivante à partir des patrons topicaux et des marqueurs aspectuels) et la *prédiction pragmatique* (anticipation du schéma narratif et du déroulé conversationnel à partir des scripts culturels). Cette *grammaire prédictive* fournit un cadre théorique pour l'apprentissage du chinois et ouvre des perspectives pour la conception de modèles linguistiquement pertinents et culturellement adaptés à l'écosystème sinophone.

Mots-clés : linguistique prédictive, chinois mandarin, grammaire prédictive, sémantique distributionnelle, syntaxe topique, pragmatique culturelle, modèles de langage, intelligence artificielle.

Abstract: This article extends the foundations of predictive linguistics (Guidère, 2015, 2025) by applying them to Mandarin Chinese. It demonstrates that the Chinese language, characterized by isolating morphology, strict word order, strong context dependence, and a dense cultural fabric, constitutes a prime area for predictive analysis. After a review of the theoretical framework derived from major language models, the article develops a *predictive grammar of Chinese* structured around three levels: *semantic prediction* (anticipation of the next word based on lexical co-occurrences and semantic radicals), *syntactic prediction* (anticipation of the next phrase and sentence based on topical patterns and aspectual markers), and *pragmatic prediction* (anticipation of the narrative structure and conversational flow based on cultural scripts). This *predictive grammar* provides a theoretical framework for learning Chinese and opens up avenues for designing linguistically relevant and culturally appropriate models for the Chinese-speaking ecosystem.

Keywords: predictive linguistics, Mandarin Chinese, predictive grammar, distributional semantics, topic syntax, cultural pragmatics, language models, artificial intelligence.

Introduction

La linguistique prédictive, telle que formulée par Guidère (2015, 2025), constitue un cadre théorique pertinent pour l'analyse du fonctionnement du langage à la lumière des architectures neuronales contemporaines. Elle repose sur une hypothèse forte : toute production langagière, qu'elle soit orale ou écrite, procède d'une anticipation mentale exercée à partir d'un contexte verbal antérieur. Dans cette perspective, le locuteur comme l'auditeur mobilisent en permanence un ensemble de régularités linguistiques qui rendent certaines continuations plus attendues que d'autres lors de la production langagière. Dans cette optique, la grammaire elle-même peut être redéfinie comme un système de règles et de contraintes qui structurent ces anticipations. Les unités concernées ne se limitent pas au mot : elles vont du trait phonologique au schéma discursif, en passant par le morphème, le syntagme et la phrase. La compétence linguistique devient ainsi, pour l'essentiel, une compétence prédictive transversal, surplombant les compétences habituelles (compréhension écrite, orale ; expression écrite, orale).

Ce paradigme a déjà fait l'objet d'applications fécondes dans plusieurs langues typologiquement distinctes (anglais, arabe, japonais, espagnol) et a révélé, pour chacune d'elles, des configurations prédictives spécifiques conditionnées par la morphologie, l'ordre des constituants et les conventions pragmatiques. Son extension au chinois mandarin, langue isolante à écriture logographique, ouvre des perspectives particulièrement stimulantes. Le chinois ne se contente pas de confirmer la pertinence du cadre général de la linguistique prédictive : il en déplace les points d'appui et en éprouve les concepts fondamentaux.

Trois notions méthodologiques structurent l'approche proposée dans le présent article.

La première est celle d'*unité de prédiction* qui, dans une langue dépourvue de morphologie flexionnelle saillante comme le chinois, ne coïncide ni avec le morphème grammatical, ni avec le mot au sens graphique occidental, mais avec le caractère (字, *zì*), dont l'organisation interne combine un radical sémantique (部首, *bùshǒu*) et une composante phonétique. La prédiction opère donc à un niveau

où forme graphique, valeur sémantique et indice phonologique sont intriqués.

La deuxième notion est celle de *fenêtre contextuelle*, que l'on formalisera à travers la probabilité conditionnelle $P(u_n | u_1 \dots u_{n-1})$ familière des modèles de type Transformer. En chinois, cette fenêtre inclut non seulement la séquence linéaire de caractères, mais aussi la structure topique-commentaire, la présence de particules aspectuelles et modales (了 *le*, 过 *guo*, 着 *zhe*, 吗 *ma*, 呢 *ne*, 吧 *ba*), ainsi que des indices culturels souvent implicites.

La troisième notion est celle de *niveau de prédiction* qui se déploie suivant les principaux niveaux d'analyse du langage, à savoir : une prédiction sémantique (anticipation de l'unité lexicale suivante), une prédiction syntaxique (anticipation du syntagme ou de la phrase suivante) et une prédiction pragmatique (anticipation du schéma narratif ou du déroulement conversationnel à venir). Ces trois niveaux ne sont pas hiérarchiquement isolés : ils s'entre-déterminent, et c'est précisément leur articulation qui fait la spécificité du chinois.

À ces concepts s'ajoutent trois traits structurels qui rendent le mandarin particulièrement instructif pour la linguistique prédictive.

D'abord, la *segmentation* y est elle-même un acte prédictif : l'absence de séparateurs orthographiques fait de l'identification des frontières de mots une tâche probabiliste, tranchée par la fréquence des cooccurrences.

Ensuite, la *tonalité* (quatre tons lexicalement distinctifs, plus un ton neutre) opère une désambiguïsation phonologique que l'écriture ne transcrit pas, mais que l'attente prédictive du locuteur intègre naturellement.

Enfin, la *contextualité* (ou degré de détermination contextuelle) atteint en chinois un seuil tel que de nombreux énoncés demeurent indéchiffrables sans la mobilisation de scripts culturels partagés, qu'il s'agisse des relations hiérarchiques, de la gestion de la face (面子, *miànzi*) ou des rituels de politesse.

Cette *contextualité* est pensée comme un instrument didactique : elle offre à l'apprenant du mandarin une manière nouvelle d'intérioriser la langue, non plus par l'accumulation de règles, mais par l'acquisition d'attentes contextuelles.

Du cadre général à la spécificité du chinois

La linguistique prédictive pose, comme hypothèse fondatrice, que la forme et le sens n'existent pas indépendamment des contextes dans lesquels ils se manifestent : ils émergent conjointement de régularités distributionnelles stabilisées par l'usage. Cette hypothèse, qui trouve ses racines dans les travaux distributionnalistes du XX^e siècle, reçoit aujourd'hui une formalisation mathématique dans les architectures neuronales contemporaines, en particulier dans les modèles de type « Transformer ». La prédiction y est exprimée sous la forme d'une probabilité conditionnelle $P(u_n / u_1 \dots u_{n-1})$, dans laquelle l'estimation d'une unité linguistique dépend d'une fenêtre contextuelle pondérée par des mécanismes d'attention. Appliqué à des langues comme l'anglais, l'arabe, le japonais ou l'espagnol, ce cadre révèle, pour chacune, des profils prédictifs spécifiques, conditionnés par la morphologie, la topologie des constituants et les conventions énonciatives. Le mandarin, toutefois, met cette mécanique à l'épreuve d'une manière singulière, au point d'exiger une reconfiguration partielle des concepts employés en linguistique prédictive.

Un premier déplacement de la théorie concerne la *segmentation*. En effet, les langues à séparateurs orthographiques, comme le français ou l'anglais, offrent au traitement automatique une première couche de découpage aisément repérable, le blanc typographique balisant les frontières lexicales. Le chinois écrit, en revanche, se présente comme un flux ininterrompu de caractères, sans marque explicite des limites de mots. La tokenisation, qu'elle opère au niveau du caractère, du mot polysyllabique ou de l'unité sous-mot, y devient elle-même une opération prédictive : la probabilité qu'une suite de caractères constitue une unité lexicale se déduit de la fréquence et de la cohésion de leurs cooccurrences. Ainsi, la chaîne 中国人 peut être segmentée en 中国|人 (« personne de Chine ») ou en 中|国人 (« compatriote central »), et seul le contexte, avec les attentes probabilistes qu'il active, permet de trancher. La segmentation n'est donc pas un préalable à la prédiction : elle est déjà, en elle-même, un acte prédictif.

Un second déplacement concerne la *phonologie*. En effet, le chinois est une langue tonale, dans laquelle quatre tons lexicalement distinctifs, auxquels s'ajoute un ton neutre, suffisent à différencier des

unités par ailleurs homographes ou homophones. Cette information tonale, absente de l'écriture courante, n'en structure pas moins les attentes du locuteur et de l'auditeur. Par exemple, les paires minimales 妈 *mā* (« mère »), 麻 *má* (« chanvre »), 马 *mǎ* (« cheval ») et 骂 *mà* (« injurier ») illustrent ce phénomène : la désambiguïsation y opère par la prédiction tonale autant que par le contexte sémantique. À cela s'ajoute une prolifération d'homophones non-tonaux, qui rend la prédiction phonologique indissociable de la prédiction sémantique : l'apprenant doit apprendre à *entendre* ce que l'écriture tait.

Un troisième déplacement, plus profond encore, concerne la *contextualité*. Certes toute langue est contextuelle, mais le chinois pousse cette propriété à un degré très avancé. L'ellipse du sujet est courante, la référence anaphorique implicite, les particules finales véhiculent des nuances illocutoires que l'intonation seule pourrait porter dans d'autres langues. Surtout, de nombreux énoncés n'accèdent à leur pleine signification que par la mobilisation de scripts culturels partagés : conscience des hiérarchies relationnelles, gestion de la face (面子, *miànzi*), codes de modestie, rituels de politesse et de réciprocité. Un simple 吃了吗? (*Chī le ma?*, littéralement « As-tu mangé? ») ne peut être interprété comme une vraie question sur l'alimentation : il fonctionne, dans la plupart des situations, comme une formule phatique équivalente à « bonjour ». Seule une anticipation pragmatique, nourrie de la connaissance des usages, permet de lever l'ambiguïté.

Ces trois traits (segmentation probabiliste, tonalité distinctive et contextualité élevée) ne sont pas des obstacles à la modélisation prédictive du chinois : ils en sont, au contraire, les ressorts les plus productifs. Ils justifient que l'on consacre au mandarin une grammaire prédictive spécifique.

Grammaire prédictive du chinois

La grammaire prédictive du chinois se distingue des grammaires descriptives traditionnelles par son objet : elle ne cherche pas à énoncer des règles closes gouvernant la combinaison des unités, mais à cartographier les régularités probabilistes qui rendent certaines continuations plus attendues que d'autres. Elle n'oppose pas « le correct » à « l'incorrect » ; elle gradue le probable, le marginal et l'inattendu. Dans cette perspective, la maîtrise du mandarin ne consiste plus à mémoriser un

ensemble de structures figées, mais à interioriser progressivement un « horizon d'attentes », structuré à trois niveaux complémentaires : celui de l'unité lexicale (sémantique), celui du syntagme et de la phrase (syntaxe), celui du discours et de l'interaction (pragmatique).

La prédiction sémantique en chinois

La prédiction sémantique concerne l'anticipation de l'unité lexicale (caractère ou mot polysyllabique) la plus probable dans un contexte donné. Elle repose sur trois mécanismes entrelacés : la *cooccurrence lexicale*, la *solidarité sémantique entre radicaux*, et les *régularités collocationnelles cristallisées* dans la langue.

Le premier mécanisme, la *cooccurrence lexicale*, est particulièrement visible dans les mots disyllabiques qui constituent l'essentiel du lexique moderne en chinois. Lorsqu'un locuteur produit le caractère 电 *diàn* (« électricité »), l'espace des continuations probables se restreint drastiquement : 电脑 *diànnǎo* (« ordinateur », littéralement « cerveau électrique »), 电话 *diànhuà* (« téléphone », litt. « parole électrique »), 电影 *diànyǐng* (« film », litt. « ombre électrique »), 电视 *diànshì* (« télévision », litt. « vision électrique »). De même, après 飞 *fēi* (« voler »), on attend 飞机 *fēijī* (« avion »), 飞行 *fēixíng* (« vol »), 飞鸟 *fēiniǎo* (« oiseau en vol »). La prédiction sémantique opère ici par appariement de caractères à forte affinité distributionnelle.

Le deuxième mécanisme tient à la *structure idéographique* elle-même. Le radical sémantique (部首, *bùshǒu*) constitue un indice prédictif puissant, qui oriente l'interprétation du caractère avant même que son contenu phonétique ne soit mobilisé. Les caractères construits sur le radical de l'eau (氵) appartiennent, avec une probabilité élevée, au champ des phénomènes liquides : 河 *hé* (« fleuve »), 海 *hǎi* (« mer »), 湖 *hú* (« lac »), 泪 *lèi* (« larme »), 汗 *hàn* (« sueur »). Ceux construits sur le radical du cœur (忄 ou 心) renvoient aux états affectifs ou mentaux : 怕 *pà* (« craindre »), 忙 *máng* (« être occupé »), 想 *xiǎng* (« penser, désirer »), 恨 *hèn* (« haïr »). Cette régularité confère au caractère inconnu une probabilité forte d'être correctement interprété sur la base du seul radical, phénomène

bien connu de l'enseignement du chinois et exploité par les modèles neuronaux travaillant à l'échelle du sous-caractère.

Le troisième mécanisme est celui des *collocations cristallisées*, souvent d'origine classique, dont la première moitié appelle la seconde avec une probabilité quasi déterministe. Ainsi, 马马 *mǎmǎ* appelle 虎虎 *hǔhǔ* pour former 马马虎虎 *mǎmǎhǔhǔ* (« comme-ci comme-ça », litt. « cheval-cheval tigre-tigre »). De même, 一石 *yī shí* (« une pierre ») appelle 二鸟 *èr niǎo* (« deux oiseaux ») dans 一石二鸟 *yī shí èr niǎo* (« faire d'une pierre deux coups »). Le *chéngyǔ* (成语), expression idiomatique en quatre caractères, est précisément l'une des formes où la prédictibilité sémantique du mandarin atteint son intensité maximale : dès lors que les deux premiers caractères sont produits, les deux suivants sont fortement contraints et aisément prédits.

Les exemples suivants illustrent différentes configurations de prédiction sémantique :

Exemple 1 — Champ sémantique contextuel : Dans l'énoncé 我渴了, 想喝一杯__ *Wǒ kě le, xiǎng hē yì bēi* __ (« J'ai soif, je voudrais boire un verre de __ »), le contexte active un champ clos : 水 *shuǐ* (« eau »), 茶 *chá* (« thé »), 咖啡 *kāfēi* (« café »), 果汁 *guǒzhī* (« jus de fruit »). D'autres continuations, lexicalement possibles, sont statistiquement improbables.

Exemple 2 — Contrainte classificatoire : Le chinois impose l'usage de spécificatifs (量词, *liàngcí*) dont la distribution est hautement prévisible à partir du nom attendu. Après 一只 *yì zhī*, on attend un animal de petite ou moyenne taille : 猫 *māo* (« chat »), 狗 *gǒu* (« chien »), 鸟 *niǎo* (« oiseau »). Après 一张 *yì zhāng*, on attend un objet plat : 纸 *zhǐ* (« feuille de papier »), 票 *piào* (« billet »), 照片 *zhàopiàn* (« photographie »), 桌子 *zhuōzi* (« table »). Le spécificatif fonctionne ainsi comme un filtre prédictif explicite.

Exemple 3 — Polarité et attente lexicale : L'énoncé 他不但聪明, 而且__ *Tā búdàn cōngmíng, érqiě* __ (« Non seulement il est intelligent, mais en plus __ ») contraint la suite à un adjectif mélioratif : 勤奋 *qínfèn* (« travailleur »), 善良 *shànliáng* (« bon »), 努力 *nǔlì* (« assidu »). Un adjectif dépréciatif,

comme 懒惰 *lǎnduò* (« paresseux »), violerait la cohérence et l'attente prédictive de la construction 不但...而且...

Exemple 4 — Prédiction par contraste : Dans 他虽然很累, 但是还是 __ *Tā suīrán hěn lèi, dànshì háishi* __ (« Bien qu'il soit fatigué, il __ quand même »), la structure concessive 虽然...但是... appelle un verbe d'effort ou de persévérance: 工作 *gōngzuò* (« travailler »), 学习 *xuéxí* (« étudier »), 坚持 *jiānchí* (« persister »).

Exemple 5 — Chényǔ : L'amorce 画蛇 *huà shé* (« dessiner un serpent ») appelle quasi exclusivement 添足 *tiān zú* (« ajouter des pattes »), pour former 画蛇添足 *huà shé tiān zú* (« en rajouter inutilement »). De même, 守株 *shǒu zhū* (« garder la souche ») projette 待兔 *dài tù* (« attendre le lièvre »), composant 守株待兔 *shǒu zhū dài tù* (« attendre sans agir que la chance revienne »).

La prédiction syntaxique en chinois

La prédiction syntaxique porte sur l'anticipation du syntagme ou de la phrase qui prolonge un énoncé en cours. Elle s'appuie, en chinois, sur un ensemble de structures corrélatives, de marqueurs aspectuels et de schémas topicaux dont la valeur prédictive est élevée, parfois supérieure à celle observée dans les langues flexionnelles. L'absence de conjugaison y est largement compensée par la rigueur de l'ordre des constituants et par la densité fonctionnelle des particules.

Le premier ressort de la *prédiction syntaxique* réside dans les *constructions corrélatives*, dans lesquelles l'apparition du premier connecteur rend l'apparition du second quasi obligatoire. La paire 因为...所以... *yīnwèi... suǒyǐ...* (« parce que... donc... ») en est l'exemple canonique: 因为下雨了, 所以我没去 *Yīnwèi xià yǔ le, suǒyǐ wǒ méi qù* (« Parce qu'il a plu, je n'y suis pas allé »). De façon analogue, 如果...就... *rúguǒ... jiù...* (« si... alors... »), 虽然...但是... *suīrán... dànshì...* (« bien que... cependant... »), 不是...而是... *búshì... érshì...* (« ce n'est pas... mais... »), 只要...就... *zhǐyào... jiù...* (« il suffit que... pour que... »), 越...越... *yuè... yuè...* (« plus... plus... ») organisent l'attente syntaxique sur une longue portée. Dès la production du premier connecteur, le locuteur projette non seulement un type de proposition, mais aussi le marqueur qui l'introduira.

Le deuxième ressort tient à la *structure topique-commentaire*, qui donne au chinois sa physionomie informationnelle. Un énoncé comme 这本书, 我已经看过了 *Zhè běn shū, wǒ yǐjīng kàn guo le* (« Ce livre, je l'ai déjà lu ») illustre la projection du topique en tête d'énoncé, suivi d'un commentaire dont la structure prédicative est, elle, très largement contrainte. Une fois le topique posé, le système anticipe un commentaire contenant un verbe transitif et ses modulations aspectuelles. Cette dissociation typologique, plus nette qu'en français, fait du topique un véritable opérateur de prédiction.

Le troisième ressort est celui des *particules aspectuelles et modales*, qui bornent les possibilités syntaxiques et sémantiques en aval. La particule 了 *le* postverbale signale l'achèvement et appelle, dans de nombreux contextes, un complément résultatif ou quantificatif: 我吃了三碗饭 *Wǒ chī le sān wǎn fàn* (« J'ai mangé trois bols de riz »). La particule 过 *guo* signale une expérience révolue: 我去过北京 *Wǒ qù guo Běijīng* (« Je suis déjà allé à Pékin »). La particule 着 *zhe* marque la continuité d'un état: 他笑着说 *Tā xiào zhe shuō* (« Il dit en souriant »). Chaque particule sélectionne une classe de continuations probables et en écarte d'autres.

Les exemples suivants illustrent différents régimes de prédiction syntaxique :

Exemple 1 — Corrélation conditionnelle : Dans 如果明天不下雨, __ *Rúguǒ míngtiān bú xià yǔ,* __ (« S'il ne pleut pas demain, __ »), la structure projette un apodose généralement introduite par 就 *jiù*: 我们就去公园 *wǒmen jiù qù gōngyuán* (« alors nous irons au parc »). L'omission de 就 est marquée et statistiquement rare.

Exemple 2 — Corrélation graduelle : L'amorce 他跑得越来越 __ *Tā pǎo de yuèlái yuè* __ (« Il court de plus en plus __ ») sélectionne un adjectif gradable: 快 *kuài* (« vite »), 慢 *màn* (« lentement »), 累 *lèi* (« fatigué »). La structure 越来越 bloque les adjectifs non gradables.

Exemple 3 — Construction 把 *bǎ* : Dans 他把门 __ *Tā bǎ mén* __ (« Il [faire subir à] la porte __ »), la construction topicalise un objet défini et projette un verbe d'action résultative: 关了 *guān le* (« a fermée »), 打开了 *dǎkāi le* (« a ouverte »), 修好了 *xiūhǎo le* (« a réparée »). Un verbe d'état (喜欢

xǐhuān, « aimer ») y est exclu.

Exemple 4 — Construction passive en 被 bèi : L'énoncé 他被老板__ *Tā bèi lǎobǎn* __ (« Il a été [par le patron] __ ») appelle un verbe transitif à polarité agentive forte: 批评了 *pīpíng le* (« critiqué »), 表扬了 *biǎoyáng le* (« félicité »), 辞退了 *cítuì le* (« licencié »).

Exemple 5 — Topique en tête : L'énoncé 中国菜, 我最喜欢的是__ *Zhōngguó cài, wǒ zuì xǐhuān de shì* __ (« La cuisine chinoise, ce que je préfère c'est __ ») appelle un nom de plat: 北京烤鸭 *Běijīng kǎoyā* (« canard laqué de Pékin »), 麻婆豆腐 *mápó dòufu* (« tofu de la grand-mère Ma »), 饺子 *jiǎozi* (« raviolis »).

Exemple 6 — Interrogation par reduplication verbale : La structure A-不-A, comme 你去不去? *Nǐ qù bu qù?* (« Tu y vas ou pas? »), projette une réponse binaire 去 *qù* / 不去 *bú qù*, et exclut les modalisations périphériques.

Exemple 7 — Série verbale : L'énoncé 他去商店__ *Tā qù shāngdiàn* __ (« Il va au magasin __ ») appelle, dans la logique des constructions à verbes en série, un verbe de finalité: 买东西 *mǎi dōngxi* (« acheter des choses »), 看朋友 *kàn péngyǒu* (« voir un ami »). Le chinois prédit ici la juxtaposition de deux prédicats reliés par une relation de but.

Exemple 8 — Négation et portée : L'amorce 我不是不__ *Wǒ bú shì bù* __ (« Ce n'est pas que je ne __ pas ») sélectionne un verbe intentionnel: 想去 *xiǎng qù* (« veux y aller »), 喜欢你 *xǐhuān nǐ* (« t'aime »). Cette double négation projette en aval une proposition adversative introduite par 只是 *zhǐshì* (« c'est seulement que... ») ou 而是 *érshì* (« c'est plutôt que... »).

La prédiction pragmatique en chinois

La prédiction pragmatique concerne les unités les plus larges: le schéma narratif, la trame conversationnelle, la séquence rituelle. À ce niveau, l'anticipation ne porte plus sur un caractère ou une proposition, mais sur le type d'acte de langage attendu, sur la forme qu'il doit prendre, et sur la position qu'il occupe dans une structure sociale. Le chinois y déploie une densité de scripts culturels

qui rend la prédiction pragmatique particulièrement structurée, davantage que dans les langues occidentales modernes.

Le premier axe de prédiction pragmatique tient aux *scripts rituels de salutation et de clôture*. L'échange 你好 *Nǐ hǎo* — 你好 *Nǐ hǎo* (« Bonjour — Bonjour ») est stable. La question 吃了吗? *Chī le ma?* (« As-tu mangé? »), formule phatique typique, n'attend pas une description du repas mais une réponse conventionnelle du type 吃了, 你呢? *Chī le, nǐ ne?* (« Oui, et toi? »). De même, au moment de la séparation, 慢走 *màn zǒu* (litt. « marche lentement », « prends soin de toi ») appelle 好的, 再见 *hǎo de, zàijiàn* (« d'accord, au revoir »). La prédiction opère ici sur la séquence appariée et ses formules stabilisées.

Le deuxième axe est celui de la *gestion de la face* (面子, *miànzi*) et de la *modestie rituelle*. À un compliment comme 你的中文说得真好! *Nǐ de Zhōngwén shuō de zhēn hǎo!* (« Tu parles vraiment bien chinois! »), la réponse culturellement attendue n'est pas un remerciement direct, mais une minoration: 哪里哪里, 还差得远 *Nǎli nǎli, hái chà de yuǎn* (« Mais non, j'ai encore beaucoup à apprendre »), ou 过奖了 *Guòjiǎng le* (« Vous m'en faites trop »). Un mot brut (谢谢 *Xièxiè*), courant en français, est marqué en contexte chinois traditionnel. De même, à l'offre 请多吃一点 *Qǐng duō chī yìdiǎn* (« Mangez donc un peu plus »), le convive répondra d'abord 不用了, 我已经吃饱了 *Búyòng le, wǒ yǐjīng chī bǎo le* (« Non merci, je suis déjà rassasié »), avant, souvent, d'accepter après une deuxième invitation.

Le troisième axe est celui des *invitations, offres et refus rituels*. Lorsqu'un hôte propose 这是一点小意思, 请收下 *Zhè shì yìdiǎn xiǎo yìsi, qǐng shōu xià* (« C'est un petit quelque chose, acceptez-le s'il vous plaît »), l'invité est censé d'abord refuser: 太客气了, 不用了 *Tài kèqi le, búyòng le* (« Vous êtes trop aimable, ce n'est pas nécessaire »). L'hôte insiste, et l'acceptation finale intervient après un ou deux refus conventionnels. Ce schéma d'*insistance-refus-acceptation* est hautement prédictible et joue un rôle structurant dans l'interaction sociale.

Le quatrième axe concerne l'*addition au restaurant*. À la fin d'un repas partagé, la séquence 我来买单

Wǒ lái mǎi dān (« C'est moi qui paie ») appelle 不行 不行, 我请客 Bù xíng bù xíng, wǒ qǐng kè (« Non non, c'est moi qui invite »). Cette compétition rituelle pour régler la note peut se dérouler sur plusieurs tours et se clôt parfois par un partage symbolique, l'enjeu étant moins la transaction monétaire que la démonstration publique de générosité.

Le cinquième axe est celui des *requêtes indirectes*. Le chinois tend à préférer des formulations atténuées, dont la forme est prédictible. Une demande directe 给我一杯水 Gěi wǒ yì bēi shuǐ (« Donne-moi un verre d'eau ») serait perçue comme brusque dans la plupart des contextes sociaux ; la forme attendue est 麻烦你, 能不能给我一杯水? Máfán nǐ, néng bu néng gěi wǒ yì bēi shuǐ ? (« Cela te dérangerait-il de me donner un verre d'eau ? »). La présence de 麻烦 máfan (« déranger ») et du couple 能不能 néng bu néng (« pouvoir ou non ») constitue un marqueur prédictif stable de la requête polie.

Le sixième axe porte sur les *schémas narratifs*. Les *chéngyǔ* sont souvent des condensés de récits dont la mention active, chez l'auditeur informé, un schéma narratif complet. Mentionner 守株待兔 shǒu zhū dài tù (« attendre le lièvre au pied de l'arbre ») convoque toute l'histoire du paysan qui, ayant vu un lièvre se briser la nuque contre une souche, abandonne son champ pour en attendre un second. De même, 塞翁失马 Sài wēng shī mǎ (« le vieux de la frontière a perdu son cheval ») évoque le récit taoïste selon lequel tout malheur peut porter un bonheur caché, et inversement. La prédiction pragmatique anticipe ici, sur la base d'une brève formule, l'ensemble d'une morale et des inférences qu'elle autorise.

Le septième axe est celui des *particules finales* qui orientent la suite attendue de la conversation. Ainsi, 吧 ba projette une proposition ou une suggestion douce et appelle une acceptation ou une nuance : 我们去吃饭吧 Wǒmen qù chī fàn ba (« Allons manger, qu'en dis-tu ? ») — 好啊 Hǎo a (« D'accord »). 呢 ne relance la question vers l'interlocuteur : 我喜欢红色, 你呢? Wǒ xǐhuān hóngsè, nǐ ne ? (« Moi j'aime le rouge, et toi ? »). 嘛 ma marque l'évidence et appelle une concession : 他是小孩嘛 Tā shì xiǎohái ma (« C'est un enfant, enfin ! »), attendant un 好吧 hǎo ba (« soit »). Chaque particule finale structure le

tour suivant.

Les exemples suivants illustrent ces *schémas prédictifs en contexte* :

Exemple 1 — Rencontre entre collègues : 张老师, 好久不见! Zhāng lǎoshī, hǎo jiǔ bú jiàn ! (« Professeur Zhang, cela fait longtemps ! ») projette 是啊, 最近怎么样? Shì a, zuìjìn zěnmeyàng ? (« Oui, comment vas-tu ces derniers temps ? »), suivi d'une phase de compliments mutuels et d'enquêtes sur la famille avant d'aborder tout sujet substantiel.

Exemple 2 — Invitation à dîner : 今晚来我家吃饭吧 Jīnwǎn lái wǒ jiā chī fàn ba (« Viens dîner chez moi ce soir ») appelle d'abord 不好意思, 太麻烦你了 Bù hǎoyìsi, tài máfan nǐ le (« Je suis confus, cela te dérangerait trop »), avant acceptation après insistance.

Exemple 3 — Don et réception d'un cadeau : Le don, accompagné de 一点小礼物, 不成敬意 Yìdiǎn xiǎo lǐwù, bù chéng jìngyì (« Un petit cadeau, bien peu de chose »), appelle 您太客气了 Nín tài kèqi le (« Vous êtes vraiment trop aimable »), puis l'acceptation, souvent sans ouverture immédiate du paquet en présence du donateur, ce qui en soi est un script prédictible.

Exemple 4 — Excuse et réparation : 对不起, 我来晚了 Duìbuqǐ, wǒ lái wǎn le (« Désolé, je suis en retard ») projette une atténuation : 没关系, 我也刚到 Méi guānxi, wǒ yě gāng dào (« Ce n'est rien, je viens juste d'arriver »), quand bien même l'interlocuteur attendrait depuis un long moment.

Exemple 5 — Clôture téléphonique : La séquence 好的, 就这样 Hǎo de, jiù zhèyàng (« Bon, on en reste là ») projette 好, 再联系 Hǎo, zài liánxì (« D'accord, on se recontacte »), avant 拜拜 báibái (« à bientôt »). Couper la communication sans cette séquence serait jugé brusque.

Exemple 6 — Négociation commerciale au marché : L'annonce du prix par le vendeur, 一百块 yì bǎi kuài (« cent yuans »), projette du côté du client 太贵了, 便宜一点吧 Tài guì le, piányi yìdiǎn ba (« C'est trop cher, faites un effort »), suivi d'une contre-offre basse, d'un refus, et d'une convergence progressive. L'absence de marchandage serait, dans ce contexte, culturellement marquée.

Exemple 7 — Demande de service entre proches :

L'amorce 跟你商量一件事 *Gēn nǐ shāngliang yí jiàn shì* (« Je voudrais discuter d'une chose avec toi ») annonce une requête dont le contenu sera différé, précédé d'une justification contextuelle longue. La demande finale, souvent introduite par 其实..., 不知道...方便不方便 *qíshí..., bù zhīdào... fāngbiàn bu fāngbiàn* (« en fait..., je ne sais pas si... te conviendrait »), est ainsi prédictible dans sa structure avant de l'être dans son contenu.

Ces trois niveaux de prédiction (sémantique, syntaxique, pragmatique) ne sont pas des strates indépendantes : ils s'entre-déterminent en permanence, chaque niveau contraignant les choix possibles aux autres niveaux. C'est dans leur articulation, plus que dans chacun d'eux pris isolément, que réside la spécificité du mandarin du point de vue prédictif.

Les modèles de langage prédictifs

Les développements qui précèdent ouvrent une perspective qui dépasse le cadre strict de la description grammaticale ou de la didactique du mandarin. En effet, la grammaire prédictive du chinois, telle que nous avons tenté de l'esquisser, n'est pas seulement un instrument pédagogique à destination des apprenants : elle constitue aussi un cadre heuristique dont les retombées concernent directement la conception des grands modèles de langage et, plus largement, les applications de traitement automatique du chinois. Cette convergence entre linguistique et intelligence artificielle mérite d'être explicitée.

Les architectures neuronales contemporaines, qu'il s'agisse des modèles de type « Transformer », des modèles de diffusion textuelle ou des systèmes hybrides, reposent toutes, en dernière analyse, sur une mécanique prédictive. Elles estiment, à chaque pas, la probabilité de l'unité suivante conditionnellement à un contexte, exactement selon la formule $P(u_n | u_1 \dots u_{n-1})$ évoquée plus haut.

Dès lors, la qualité d'un modèle dépend étroitement de l'adéquation entre les régularités qu'il a su apprendre et celles qui gouvernent effectivement la langue cible. Or les grands modèles existants, majoritairement entraînés sur des corpus dominés par l'anglais, tendent à traiter le chinois comme une langue parmi d'autres, sans prise en compte

systématique des trois traits structurels que nous avons identifiés : la segmentation probabiliste, la tonalité distinctive et la contextualité culturelle élevée.

Cette homogénéisation produit des modèles qui, bien qu'opérationnels, reproduisent mal les profils prédictifs propres au mandarin et génèrent, en production, des énoncés formellement corrects mais pragmatiquement décalés.

La première contribution d'une *grammaire prédictive du chinois* à la conception de tels modèles concerne la *tokenisation*. Une segmentation aveugle en sous-mots de type « Byte Pair Encoding », héritée du traitement de l'anglais, fragmente arbitrairement les caractères composés et brise la solidarité radical-phonétique qui structure l'idéogramme. Une *tokenisation informée par la linguistique prédictive* intégrerait, à l'inverse, la dimension du radical (部首), la cohésion des mots disyllabiques et la typicité des *chéngyǔ*. Elle reconnaîtrait que la probabilité de la séquence 电脑 n'est pas réductible au produit des probabilités isolées de 电 et de 脑, mais qu'elle exprime une *unité lexicale cristallisée* dont la cohésion est elle-même porteuse d'information.

La deuxième contribution concerne le *traitement des particules et des structures corrélatives*. Les particules aspectuelles (了, 过, 着), modales (吗, 呢, 吧) et les paires corrélatives (因为...所以..., 虽然...但是..., 不但...而且...) sont des opérateurs prédictifs à longue portée, dont la fonction dépasse la simple marque syntaxique. Un *modèle prédictif adapté au chinois* devrait reconnaître explicitement ces marqueurs comme des points de branchement, à partir desquels l'espace des continuations probables se restructure. Plutôt que de laisser l'apprentissage statistique inférer, à partir de masses de données, ce que la langue marque de façon explicite, une *architecture linguistiquement informée* intégrerait ces contraintes comme prérequis structuraux (priors), à la manière de biais inductifs.

La troisième contribution, la plus riche de conséquences, concerne la *dimension pragmatique et culturelle*. Un modèle de langage qui produit, en réponse au compliment 你的中文说得真好, un simple 谢谢 peut être techniquement correct mais

culturellement inadéquat. Il manque au locuteur chinois l'attente d'une minoration rituelle (哪里哪里, 还差得远). De même, un modèle qui accepte une invitation dès la première offre, sans passer par la séquence conventionnelle de refus rituel, rompt un script social stabilisé.

Ces décalages, parfois imperceptibles aux évaluations quantitatives classiques, grèvent lourdement l'acceptabilité sociale des productions en contexte sinophone. Une *grammaire prédictive pragmatique* devrait fournir au concepteur de modèle un *répertoire explicite des schémas conversationnels attendus*, qui peut nourrir aussi bien les données d'entraînement que les fonctions d'évaluation et les stratégies d'alignement.

Cette exigence d'*adaptation culturelle* ne doit pas être confondue avec une simple localisation de surface, qui consisterait à traduire les interfaces ou à ajuster quelques préférences stylistiques. Il s'agit, plus profondément, de reconnaître que la langue n'est jamais un code culturellement neutre, mais un système d'attentes partagées qui encode, dans sa structure même, des représentations du monde, des hiérarchies sociales et des régimes d'interaction. En effet, la gestion de la face (面子), le jeu des rapports entre *lǐ* (礼, « rite ») et *rén* (仁, « humanité »), la valorisation de la modestie, la préférence pour l'implicite et l'indirection, autant de traits qui ne sont pas des ornements pragmatiques surajoutés à une grammaire neutre, mais des composantes intrinsèques de la prédictibilité du chinois. Un *modèle culturellement adapté* est un modèle qui a appris à anticiper non seulement ce qui est dit, mais aussi ce qui, dans un contexte donné, doit être tu, différé ou indiqué par la seule forme de l'énoncé.

Sur un plan plus prospectif, l'articulation entre linguistique prédictive et intelligence artificielle dessine les contours d'une nouvelle génération de modèles que l'on pourrait qualifier de *linguistiquement pertinents*.

Pertinents, d'abord, parce qu'ils reconnaîtraient la spécificité typologique des langues qu'ils traitent, au lieu de les ramener à un patron commun dicté par les langues majoritaires des corpus d'entraînement.

Pertinents, ensuite, parce qu'ils intégreraient, à côté de l'apprentissage distributionnel massif, des connaissances linguistiques explicites (radicaux, spécificatifs, particules, corrélatifs, schémas pragmatiques) susceptibles d'être mobilisées

comme biais inductifs ou comme modules de contrôle.

Pertinents, enfin, parce qu'ils pourraient s'évaluer sur des critères non plus seulement formels, mais culturellement situés : adéquation pragmatique, respect des scripts rituels, maîtrise de la modulation indirecte.

Une telle orientation aurait des conséquences qui dépassent la seule qualité technique des systèmes. Elle contribuerait à réduire l'hégémonie linguistique et culturelle des modèles anglo-centrés, à préserver la diversité des profils prédictifs propres à chaque langue, et à faire de l'intelligence artificielle un instrument de médiation culturelle plutôt qu'un vecteur d'uniformisation.

Le chinois, par la radicalité de sa typologie et la densité de ses scripts culturels, constitue à cet égard un cas d'école : ce que la linguistique prédictive parvient à formaliser de la spécificité du mandarin peut, *mutatis mutandis*, être appliqué à toute langue non indo-européenne dont la richesse est aujourd'hui insuffisamment captée par les systèmes dominants.

Conclusion

La linguistique prédictive, dans les développements qui précèdent, s'est révélée être bien davantage qu'un simple cadre d'analyse statistique : elle fournit une matrice théorique féconde pour repenser la grammaire du chinois mandarin. Là où les approches descriptives traditionnelles procédaient par catalogue de règles et inventaire de structures, la perspective prédictive invite à déplacer le regard vers ce qui rend une continuation probable, vers ce qui organise l'attente du locuteur et de l'auditeur, vers ce qui fait qu'une langue n'est jamais un ensemble infini de combinaisons possibles, mais un espace hiérarchisé de probabilités contextuelles. En cela, la linguistique prédictive ne se substitue pas aux approches antérieures : elle les refonde en leur conférant un socle probabiliste explicite, qui rend compte aussi bien des régularités massives que des variations fines observées dans l'usage.

L'articulation des trois niveaux d'analyse (sémantique, syntaxique et pragmatique) que nous avons successivement parcourus n'est pas une simple commodité exposition. Elle reflète l'architecture même de la compétence linguistique en chinois.

Au niveau sémantique, nous avons vu la prédiction opérer à l'échelle la plus fine, celle du caractère et du mot, à travers la densité des collocations disyllabiques, la productivité des radicaux graphiques, la cristallisation des *chéngyǔ* et la contrainte exercée par les classificateurs.

Au niveau syntaxique, la prédiction s'est déployée à l'échelle du syntagme et de la phrase, mobilisant les amorces structurales comme 把 *bǎ* et 被 *bèi*, les paires corrélatives telles que 因为...所以..., 虽然...但是..., 不但...而且..., les particules aspectuelles 了, 过, 着, et la topique-commentaire comme schéma informationnel dominant.

Au niveau pragmatique enfin, la prédiction a pris la forme d'une anticipation de séquences rituelles, d'actes de langage attendus et de schémas narratifs condensés, où se manifestent la gestion de la face (面子, *miànzi*), les stratégies de refus indirect, les jeux de modestie conventionnelle et les figures classiques du déroulé discursif comme la structure *qǐ-chéng-zhuǎn-hé* (起承转合).

Ces trois niveaux, loin d'être étanches, s'entredéterminent en permanence : le choix d'une particule finale suppose une anticipation pragmatique, de même que la reconnaissance d'un topique préfigure une structure syntaxique, et l'identification d'un radical oriente une prédiction sémantique.

Il importe cependant de souligner que les ressorts prédictifs mis en évidence ne constituent en aucune façon des survivances ornementales, des curiosités typologiques ou des raffinements optionnels. Ils forment, au contraire, le cœur opératoire de la langue et le socle de son intelligibilité. Là où une grammaire descriptive pourrait traiter les particules modales comme des éléments périphériques, la linguistique prédictive les replace au centre : elles sont des opérateurs d'orientation discursive dont la portée prédictive s'étend parfois sur plusieurs tours de parole. Là où les scripts culturels pourraient passer pour une *glose* pragmatique surajoutée à une grammaire neutre, ils apparaissent, dans notre cadre, comme des composantes constitutives de la compétence linguistique elle-même : un locuteur du mandarin qui ignorerait le script du refus rituel ou la convention de la modestie face au compliment serait, littéralement, incapable de prédire

correctement la suite de ses échanges, et donc d'y participer pleinement. En ce sens, la linguistique prédictive confirme une intuition ancienne, selon laquelle il n'est point de maîtrise du chinois qui ne soit, indissociablement, maîtrise d'une culture.

La grammaire prédictive que nous avons esquissée s'adresse, nous l'avons dit, à un double public. Pour l'apprenant du mandarin, elle offre une méthode d'appropriation progressive et contextualisée, qui rompt avec la traditionnelle accumulation de règles et de listes. Plutôt que d'apprendre une grammaire comme un ensemble de prescriptions à mémoriser, l'apprenant est invité à développer une sensibilité aux régularités, à intérioriser des horizons d'attentes, à s'entraîner à anticiper ce qui vient (mot, syntagme, proposition, acte de langage).

Cette orientation pédagogique rejoint d'ailleurs les acquis les plus récents des sciences cognitives du langage, qui soulignent le rôle central de l'anticipation dans la compréhension et la production langagières, et qui mettent en évidence la supériorité des approches contextualisées sur les approches purement règlementaires.

Elle rejoint également les pratiques traditionnelles les plus éprouvées de l'enseignement du chinois, qui ont toujours valorisé la mémorisation de phrases entières, de poèmes et de textes classiques, précisément parce que ces formes condensent, dans leur tissu même, un trésor de régularités prédictives que l'apprenant absorbe sans toujours pouvoir les formuler.

Il y a là, pour l'apprenant, une manière renouvelée d'accéder à la langue : non plus par la mémorisation de règles closes, mais par l'acquisition progressive d'un horizon d'attentes, sémantiques, syntaxiques et pragmatiques, dont l'intériorisation constitue la véritable compétence.

Pour le concepteur de modèles de langage, la grammaire prédictive offre une ressource d'une autre nature : une cartographie explicite des régularités à modéliser pour produire une intelligence artificielle sinophone linguistiquement pertinente et culturellement adaptée. Les grands modèles actuels, entraînés massivement sur des corpus hétéroclites, découvrent empiriquement une partie de ces régularités, mais ils le font de manière opaque, sans garantie de complétude ni de stabilité, et au prix d'une consommation de données et de ressources considérable. Une grammaire

prédictive dédiée permet, à l'inverse, d'injecter dans l'architecture des biais inductifs ciblés, d'orienter l'apprentissage vers les configurations culturellement saillantes, d'évaluer les productions non plus seulement sur des critères de plausibilité statistique mais sur des critères d'adéquation pragmatique. Elle ouvre ainsi la voie à des modèles plus économes, plus interprétables et, surtout, plus respectueux de la singularité linguistique et culturelle du mandarin.

En ce sens, la grammaire prédictive du chinois que nous avons esquissée n'est pas un aboutissement, mais un commencement. Elle appelle des développements empiriques : constitution de corpus annotés pour chaque niveau prédictif, évaluation systématique des modèles sur des critères pragmatiques, élaboration de ressources didactiques adaptées, etc. Elle appelle également des prolongements théoriques : articulation avec les sciences cognitives, avec la sémiotique culturelle, avec l'histoire de la langue. Elle dessine, surtout, le programme d'une linguistique appliquée d'un type nouveau, dans laquelle l'enseignement des langues et la conception des intelligences artificielles cesseraient d'être deux 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. Le chinois, par le défi qu'il lance aux paradigmes dominants, n'est pas seulement un objet d'étude parmi d'autres : il est, dans le champ de la linguistique prédictive, un laboratoire privilégié, où se joue la possibilité même d'une intelligence artificielle véritablement plurilingue et culturellement sensible.

Références

- Chen, T., & Zhang, D. (2023). Different aspects of vocabulary depth knowledge in L2 Chinese reading comprehension: Comparing higher-and lower-proficiency readers. *Foreign Language Annals*, 56(3), 786-806.
- Chen, X., Wang, S., & Hartsuiker, R. J. (2022). Error-based structure prediction in language comprehension: Evidence from verb bias effects in a visual-world structural priming paradigm for Mandarin Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 48(1), 60.
- Chen, Y., Li, L., Wang, M., & Wang, R. (2022). Which cognitive factors predict L2 grammar learning: Cognitive control, statistical learning, working memory, or attention?. *Frontiers in psychology*, 13, 943988.
- Fang, H., Jiang, G., & Li, D. (2022). Sentiment analysis based on Chinese BERT and fused deep neural networks for sentence-level Chinese e-commerce product reviews. *Systems Science & Control Engineering*, 10(1), 802-810.
- 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, JARHSS 2025. Vol. 1. N°1. 72-89. ISSN 2677-7193.
- Hao, Y., Wang, X., Bin, S., Yang, Q., & Liu, H. (2024). How syntactic complexity indices predict Chinese L2 writing quality: An analysis of unified dependency syntactically-annotated corpus. *Assessing Writing*, 61, 100847.
- Huang, Z., Feng, C., & Qu, Q. (2023). Predicting coarse-grained semantic features in language comprehension: evidence from ERP representational similarity analysis and Chinese classifier. *Cerebral Cortex*, 33(13), 8312-8320.
- Li, L., Li, J., Wang, H., & Nie, J. (2024). Application of the transformer model algorithm in chinese word sense disambiguation: a case study in chinese language. *Scientific reports*, 14(1), 6320.
- Li, Y., Qin, S., Ye, J., Huang, H., Li, Y., Guo, S. Y., ... & Yu, P. S. (2025, July). Rethinking the roles of large language models in chinese grammatical error correction. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 6: Industry Track)* (pp. 553-567).
- Liu, G. L. (2024). Modeling L2 motivation change and its predictive effects on learning behaviors in the extramural digital context: A quantitative investigation in China. *Linguistics Vanguard*, 10(1), 335-343.
- Liu, M., & Dong, L. (2023). A longitudinal study of foreign language reading anxiety and its predicting effects on Chinese university students' reading English performance. *The Asia-Pacific Education Researcher*, 32(6), 783-791.
- Lu, X., & Wu, J. (2022). Noun-phrase complexity measures in Chinese and their relationship to L2 Chinese writing quality: A comparison with topic-comment-unit-based measures. *The Modern*

Language Journal, 106(1), 267-283.

Peng, Y., Sun, J., Quan, J., Wang, Y., Lv, C., & Zhang, H. (2023). Predicting Chinese EFL learners' human-rated writing quality in argumentative writing through multidimensional computational indices of lexical complexity. *Assessing Writing*, 56, 100722.

Qin, C., Zhang, R., & Xiao, Y. (2022). A questionnaire-based validation of metacognitive strategies in writing and their predictive effects on the writing performance of English as foreign language student writers. *Frontiers in Psychology*, 13, 1071907.

So, W. C., & Song, X. K. (2023). Whose gestures are more predictive of expressive language abilities among Chinese-speaking children with autism? A comparison of caregivers' and children's gestures. *Journal of Autism and Developmental Disorders*, 53(9), 3449-3459.

Sun, P. P. (2023). The influence of cognitive, affective, and sociocultural individual differences on L2 Chinese speech performance: a multiple indicators multiple causes approach. *Journal of Psycholinguistic Research*, 52(1), 217-239.

Wang, Q., Liu, W., Wang, X., Chen, X., Chen, G., & Wu, Q. (2022). A spatial-temporal graph model for pronunciation feature prediction of Chinese poetry. *IEEE Transactions on Neural Networks and Learning Systems*, 34(12), 10294-10308.

Wei, P., Wang, X., & Dong, H. (2023). The impact of automated writing evaluation on second language writing skills of Chinese EFL learners: A randomized controlled trial. *Frontiers in Psychology*, 14, 1249991.

Wei, W., Huang, Z., Feng, C., & Qu, Q. (2023). Predicting phonological information in language comprehension: evidence from ERP representational similarity analysis and Chinese idioms. *Cerebral Cortex*, 33(15), 9367-9375.

Wu, X., & Wu, Y. (2022, December). From spelling to grammar: A new framework for Chinese grammatical error correction. In *Findings of the Association for Computational Linguistics: EMNLP 2022* (pp. 889-902).

Xu, M., Wei, Z., & Wu, J. (2022). How emotional communication happens in social media: Predicting "Arousal-Homophily-Echo" emotional communication with multi-dimensional

features. *Telematics and Informatics Reports*, 8, 100019.

Xu, X., Ji, C., Li, T., & Pickering, M. J. (2025). The prediction of segmental and tonal information in Mandarin Chinese: An eye-tracking investigation. *Language, Cognition and Neuroscience*, 40(1), 56-70.

Yang, H., He, Y., Bu, X., Xu, H., & Guo, W. (2023). Automatic essay evaluation technologies in Chinese writing—a systematic literature review. *Applied Sciences*, 13(19), 10737.

Ye, J., Li, Y., Ma, S., Xie, R., Wu, W., & Zheng, H. T. (2022). Focus is what you need for Chinese grammatical error correction. *arXiv preprint arXiv:2210.12692*.

Zhang, J., & Zhang, L. J. (2023). Examining the relationship between English as a foreign language learners' cognitive abilities and L2 grit in predicting their writing performance. *Learning and Instruction*, 88, 101808.

Zhang, R., & Chen, Y. (2026). What can multi-factors contribute to Chinese EFL learners' implicit L2 knowledge?. *International Review of Applied Linguistics in Language Teaching*, 64(1), 173-195.

Zhao, X., Lan, G., & Zhang, H. (2024). The predictive effect of language achievement on multiple emotions in languages other than English: Validating a distal mediation model based on the control-value theory. *Applied Linguistics Review*, 15(6), 2713-2732.

Zhao, Y., & Wu, X. (2022). Predicting reading fluency in Chinese deaf and hard of hearing students: Contributions of character recognition, expressive vocabulary, and syntactic awareness. *American Annals of the Deaf*, 166(5), 663-680.

Zheng, R., Zhang, M., Guo, T., Guasch, M., & Ferré, P. (2023). Emotional prototypicality ratings for 636 Chinese words: A database of Chinese words with affective information. *Journal of Psycholinguistic Research*, 52(6), 2775-2792.

Zhou, W., & Gao, B. (2025). Construction and application of English-Chinese multimodal emotional corpus based on artificial intelligence. *International Journal of Human-Computer Interaction*, 41(3), 1782-1793.