

LA PRÉDICTION LINGUISTIQUE EN ARABE : DE LA THÉORIE À LA PRATIQUE

Linguistic Prediction in Arabic: From Theory to Practice

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

Chirine CHAMSINE. Professeure à l'Université du Québec à Montréal (UQÀM, Canada)

GRÉLIA (Groupe de Recherche et d'Étude en Langues et Intelligence Artificielle)

chamsine.chirine@uqam.ca

Résumé : Cet article s'inscrit dans le prolongement des fondements de la linguistique prédictive (Guidère, 2015, 2025) en les appliquant à la langue arabe. Il démontre que l'arabe, caractérisé par une morphologie dérivationnelle d'une remarquable productivité, un système de racines trilitères, une flexion casuelle et un appareil rhétorique et culturel d'une grande densité, constitue un terrain particulièrement fécond pour l'analyse prédictive. Après un rappel du cadre théorique issu des grands modèles de langage, l'article développe une *grammaire prédictive de l'arabe* articulée autour de trois niveaux : la *prédiction sémantique* (anticipation de l'unité lexicale suivante à partir des schèmes dérivationnels et des solidarités racinaires), la *prédiction syntaxique* (anticipation du syntagme et de la phrase suivante à partir des patrons de régime verbal, de la flexion casuelle et des connecteurs logiques) et la *prédiction pragmatique* (anticipation du schéma discursif et du déroulé conversationnel à partir des conventions rhétoriques et des scripts culturels arabo-islamiques). Cette *grammaire prédictive* fournit un cadre théorique pour l'apprentissage de l'arabe et ouvre des perspectives pour la conception de modèles linguistiquement pertinents et culturellement adaptés à l'écosystème arabophone.

Mots-clés : linguistique prédictive, arabe, grammaire prédictive, morphologie dérivationnelle, racines trilitères, flexion casuelle, pragmatique culturelle, modèles de langage, intelligence artificielle.

Abstract: This article extends the foundations of predictive linguistics (Guidère, 2015, 2025) by applying them to the Arabic language. It demonstrates that Arabic, characterized by a remarkably productive derivational morphology, a trilateral root system, case inflection, and a dense rhetorical and cultural fabric, constitutes a particularly fertile ground for predictive analysis. After a review of the theoretical framework derived from major language models, the article develops a predictive grammar of Arabic structured around three levels: semantic prediction (anticipation of the next lexical unit based on derivational patterns and root solidarities), syntactic prediction (anticipation of the next phrase and sentence based on verbal government patterns, case inflection and logical connectors), and pragmatic prediction (anticipation of the discursive structure and conversational flow based on rhetorical conventions and Arab-Islamic cultural scripts). This predictive grammar provides a theoretical framework for learning Arabic and opens up avenues for designing linguistically relevant and culturally appropriate models for the Arabic-speaking ecosystem.

Keywords: predictive linguistics, Arabic, predictive grammar, derivational morphology, trilateral roots, case inflection, cultural pragmatics, language models, artificial intelligence.

Introduction

La prédiction linguistique désigne le mécanisme psycholinguistique par lequel le système cognitif humain anticipe activement les unités langagières à venir en s'appuyant sur le contexte sémantique, syntaxique et pragmatique. Plutôt que de traiter l'information de manière purement réactive, le cerveau mobilise ses connaissances antérieures et les probabilités statistiques pour pré-activer des représentations mentales, ce qui optimise la fluidité et la rapidité de la compréhension. Ainsi, la prédiction constitue un pilier essentiel de l'efficacité communicationnelle, faisant écho aux architectures computationnelles contemporaines qui structurent les modèles de langage artificiels.

Dans cette perspective, l'émetteur et le récepteur ne sont plus envisagés comme deux instances distinctes aux fonctions asymétriques, mais comme deux pôles engagés dans une même activité prédictive. L'un produit en anticipant les attentes de l'autre, tandis que l'autre interprète en formulant des hypothèses sur la suite du discours. Cette convergence repose sur l'existence de régularités linguistiques partagées, qui opèrent comme des schémas probabilistes internalisés. Ces régularités ne se limitent pas à la syntaxe ou au lexique, mais traversent l'ensemble des niveaux du langage. La grammaire elle-même ne constitue plus un ensemble statique de règles et de normes, mais un système de contraintes hiérarchisées qui orientent les anticipations, depuis les traits phonologiques jusqu'aux structures discursives complexes.

Dans ce cadre, la compétence linguistique ne peut plus être réduite à la seule capacité de produire ou de comprendre des énoncés. Elle correspond avant tout à une aptitude à anticiper, à ajuster en permanence des hypothèses sur la base d'indices partiels et évolutifs. Cette compétence prédictive traverse et englobe les compétences traditionnellement distinguées, en les intégrant dans une dynamique unifiée où perception et production relèvent d'un même mécanisme d'inférence.

L'intérêt de ce paradigme tient également à sa portée typologique. Des langues aussi différentes que l'anglais, le chinois, le japonais ou le français confirment la validité de cette approche, tout en mettant en évidence des modalités prédictives spécifiques. Ces variations tiennent notamment à la morphologie, à l'ordre des constituants et aux conventions pragmatiques propres à chaque système linguistique.

L'extension de cette approche à l'arabe introduit toutefois un changement d'échelle analytique. En raison de sa morphologie dérivationnelle fondée sur des racines consonantiques et des schèmes vocaliques, l'arabe ne se contente pas d'illustrer le modèle prédictif, il en met à l'épreuve les catégories.

Trois notions méthodologiques permettent de structurer cette approche dans le cas de l'arabe.

La première est celle d'*unité de prédiction*. En effet, dans les langues indo-européennes, cette unité coïncide souvent avec le mot ou le morphème. En arabe, une telle correspondance devient inadéquate. L'unité pertinente se situe à un niveau plus abstrait, celui de la racine trilitère (جذر), qui constitue le noyau consonantique porteur d'un champ sémantique. L'insertion de cette racine dans un schème dérivationnel, le *wazn* (وزن), produit une série de formes lexicales liées entre elles par des régularités morphologiques et sémantiques. La prédiction ne s'effectue donc pas à partir de mots isolés, mais à partir d'un réseau structuré où racine, schème et valeur sémantique interagissent. Anticiper un énoncé en arabe revient ainsi, dans une large mesure, à projeter des configurations morphologiques possibles à partir d'un noyau consonantique donné.

La deuxième notion est celle de *fenêtre contextuelle*. Celle-ci peut être formalisée à l'aide de la probabilité conditionnelle, bien connue dans les modèles de langage, mais dans le cas de l'arabe, cette fenêtre excède largement la simple linéarité de la chaîne verbale. Elle intègre des informations morphologiques fines, telles que les désinences casuelles relevant de l'*i'rāb* (إعراب), qui signalent les fonctions syntaxiques, ainsi que des particules à forte valeur opératoire comme *inna* (إِنَّ), *lam* (لَمْ), *lan* (لَنْ) ou *anna* (أَنَّ), qui modulent l'assertion, la négation ou la subordination. À ces indices formels s'ajoutent des paramètres pragmatiques et culturels souvent implicites, qui orientent fortement les attentes interprétatives. La fenêtre contextuelle apparaît dès lors comme un espace multidimensionnel où se combinent des indices de nature hétérogène.

La troisième notion est celle de *niveau de prédiction*. En effet, l'anticipation linguistique ne se limite pas à un seul plan d'analyse, mais se déploie simultanément à plusieurs niveaux. Au *niveau sémantique*, elle concerne l'identification probable de l'unité lexicale à venir en fonction du champ notionnel activé. Au *niveau syntaxique*, elle porte sur la structuration attendue des constituants, qu'il s'agisse d'un syntagme ou d'une phrase complète. Au *niveau*

pragmatique enfin, elle implique une projection du déroulement discursif, qu'il prenne la forme d'un récit, d'une argumentation ou d'un échange conversationnel. Ces différents niveaux ne fonctionnent pas de manière indépendante, mais s'articulent dans un système intégré où chaque indice contribue à affiner les hypothèses sur la suite du discours.

Ces trois niveaux ne sont pas hiérarchiquement cloisonnés : ils s'entre-déterminent, et c'est précisément leur articulation qui fait la spécificité de l'arabe.

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

D'abord, la *morphologie dérivationnelle* y est elle-même un moteur prédictif : le système des racines trilitères, combiné aux schèmes dérivationnels (*awzān*), permet de générer, à partir d'un noyau consonantique donné, un réseau de mots apparentés dont la cohésion sémantique oriente fortement l'anticipation.

Ensuite, la *flexion casuelle* (*i'rāb*), héritée de l'arabe classique et partiellement maintenue dans le registre soutenu moderne, fournit des indices prédictifs syntaxiques d'une précision remarquable : la terminaison d'un mot signale sa fonction grammaticale et contraint la nature du constituant suivant.

Enfin, la *contextualité culturelle*, issue de l'ancrage dans une tradition rhétorique, est adossée sur une *diglossie* dans laquelle coexistent un registre standard normé et des dialectes régionaux aux traits aussi prédictifs. Dans cet environnement linguistique et culturel, de nombreux énoncés ne prennent leur pleine signification que par la mobilisation de scripts partagés, qu'il s'agisse de formules religieuses, de conventions d'hospitalité ou de codes d'honneur.

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

La spécificité prédictive de la langue arabe

Dans le cas de l'arabe, la prédiction mobilise non seulement les segments précédents de la chaîne verbale, mais aussi des contraintes dérivationnelles, des indices flexionnels et des cadres rhétoriques souvent implicites. La relation entre les unités ne se limite pas à leur ordre d'apparition, mais s'inscrit dans

un réseau de dépendances plus profond, où la racine, le schème et les marques grammaticales contribuent conjointement à orienter les attentes.

Ainsi, la spécificité prédictive de l'arabe ne réside pas simplement dans une variation typologique parmi d'autres. Elle engage une reconfiguration partielle des outils analytiques de la linguistique prédictive. Elle invite à dépasser une conception strictement séquentielle de la prédiction pour intégrer des formes de structuration non linéaires, où l'anticipation linguistique se déploie dans un espace multidimensionnel articulant morphologie, syntaxe et pragmatique.

Un premier déplacement de la théorie concerne la *morphologie dérivationnelle*. L'arabe repose sur un système de racines consonantiques, majoritairement trilitères, qui constituent le noyau sémantique stable autour duquel gravitent les formes dérivées. La racine *كتب* (k-t-b), par exemple, véhicule l'idée générale d'écriture et engendre, par insertion dans des schèmes vocaliques distincts, des dizaines de dérivés : *كَتَبَ kataba* (« il a écrit »), *كِتَاب kitāb* (« livre »), *كَاتِب kātib* (« écrivain »), *مَكْتَبَة maktaba* (« bibliothèque »), *مَكْتُوب maktūb* (« écrit, lettre »). Ce mécanisme fait de la dérivation morphologique un acte prédictif en soi : la reconnaissance de la racine active immédiatement un réseau de formes attendues et la production de l'une de ces formes projette les autres dans le champ des continuations possibles.

Un second déplacement concerne la *flexion casuelle*. À la différence des langues isolantes comme le chinois ou des langues à morphologie réduite comme l'anglais contemporain, l'arabe classique et l'arabe standard moderne conservent un système de désinences casuelles (*إعراب, i'rāb*) qui marque la fonction grammaticale de chaque nom dans la phrase : le nominatif (*رَفْع, raf'*) pour le sujet, l'accusatif (*نَصْب, naṣb*) pour le complément d'objet, le génitif (*جَر, jarr*) après une préposition. Ces marques casuelles fonctionnent comme des opérateurs prédictifs d'une précision remarquable : elles signalent la fonction du constituant en cours et contraignent la nature de ce qui suit. Ainsi, un nom au génitif (cas indirect) après une préposition projette un complément circonstanciel ou un argument locatif, tandis qu'un nom à l'accusatif (cas direct) post-verbal signale un complément d'objet et oriente l'attente vers une suite prédictive.

Un troisième déplacement, plus large encore, concerne la *diglossie* et la *contextualité culturelle*. L'arabe vit dans une tension permanente entre un registre littéraire normé (*فصحى, fuṣḥā*) et des variétés

dialectales régionales (عامية, *‘āmmiyya / dārija*) qui divergent considérablement sur les plans phonologique, lexical et syntaxique. Cette coexistence impose au locuteur un choix de registre qui est lui-même un acte prédictif tributaire du contexte (situation formelle, médias, conversation familiale) qui détermine le registre de langue, lequel contraint l'ensemble des formulations disponibles.

À cette diglossie s'ajoute un enracinement dans une tradition rhétorique et religieuse d'une densité considérable. Les formules coraniques, les invocations rituelles, les proverbes et les structures rhétoriques héritées de la *balāgha* (بلاغة, « rhétorique ») structurent l'énonciation quotidienne à un degré rarement atteint dans d'autres langues. Un simple بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ (*Bismi llāhi r-rahmāni r-rahīm*, Au nom de Dieu, le Clément, le Miséricordieux) ne fonctionne pas comme une prière isolée : c'est un marqueur pragmatique d'ouverture qui projette une action à venir (discours, voyage, réception, repas, etc.) et qui appelle, en réponse, des formules spécifiques déterminées par le contexte.

Ces trois traits (morphologie dérivationnelle, flexion casuelle et contextualité culturelle) ne sont pas des obstacles à la modélisation prédictive de l'arabe : ils en sont, au contraire, les ressorts les plus productifs et justifient que l'on consacre à l'arabe une grammaire prédictive qui lui est spécifique.

Grammaire prédictive de l'arabe

La grammaire prédictive de l'arabe se démarque des approches descriptives classiques par un déplacement de son objet d'analyse. Elle ne cherche pas à formuler un ensemble de règles fixes régissant la combinaison des unités, mais à identifier et modéliser les régularités probabilistes qui orientent les enchaînements possibles du discours. Dans cette optique, l'opposition entre « correct » et « incorrect » cède la place à une évaluation graduelle des productions, allant du hautement attendu au marginal, voire à l'imprévisible.

Une telle conception transforme en profondeur la notion de « maîtrise linguistique ». Il ne s'agit plus d'accumuler des paradigmes stabilisés, mais de développer une sensibilité aux configurations les plus plausibles dans un contexte donné. Cette compétence repose sur l'intégration progressive d'un horizon d'attentes structuré, qui s'organise selon trois plans étroitement articulés. Le premier concerne le niveau lexical et sémantique, où s'élabore l'anticipation des unités de sens.

Le deuxième relève du niveau syntaxique, impliquant la projection des structures phrastiques et des relations entre constituants. Le troisième, enfin, engage le niveau pragmatique, où se dessinent les dynamiques discursives et interactionnelles qui orientent l'interprétation globale.

La prédiction sémantique en arabe

La prédiction sémantique concerne l'anticipation de l'unité lexicale la plus probable dans un contexte donné. Elle repose, en arabe, sur trois mécanismes entrelacés : la *productivité des racines consonantiques*, la *contrainte des schèmes dérivationnels*, et les *régularités collocationnelles* inscrites dans la langue.

Le premier mécanisme, la *productivité racinaire*, est le trait le plus distinctif de l'arabe du point de vue prédictif. Lorsqu'un locuteur produit un mot construit sur la racine علم ('a-l-m), liée à l'idée de savoir, l'espace des continuations sémantiques se structure autour d'un réseau cohérent : عِلْم *'ilm* (« science »), عالِم *'ālim* (« savant »), معلومة *ma'lūma* (« information »), تَعْلِيم *ta'līm* (« enseignement »), مُعَلِّم *mu'allim* (« enseignant »). De même, la racine حَكَم (ḥ-k-m) véhicule l'idée de jugement et de gouvernement : حُكْم *ḥukm* (« jugement »), حَكِيم *ḥakīm* (« sage »), حِكْمَةٌ *ḥikma* (« sagesse »), مَحْكَمَةٌ *maḥkama* (« tribunal »), حُكُومَةٌ *ḥukūma* (« gouvernement »). La prédiction sémantique opère ici par activation d'un réseau racinaire à forte cohésion distributionnelle.

Le deuxième mécanisme tient aux *schèmes dérivationnels* (أوزان, *awzān*) qui constituent un indice prédictif majeur. Le schème فاعِل (*fā'il*) désigne l'agent avec une régularité élevée : كاتب *kātib* (« écrivain »), عامل *āmil* (« ouvrier »), حاكم *ḥākim* (« gouverneur »). Le schème مَفْعَل (*maf'al*) désigne le lieu : مَكْتَب *maktab* (« bureau »), مَلْعَب *mal'ab* (« terrain de jeu »), مَطْبَخ *maṭbaḥ* (« cuisine »). Le schème فاعِل (*fā'il*) porte souvent un sens intensif ou permanent : كبير *kabīr* (« grand »), جميل *jamīl* (« beau »), قديم *qadīm* (« ancien »). Cette régularité confère au schème une puissance de prédiction catégorielle : la reconnaissance du patron morphologique oriente immédiatement l'interprétation du mot, même inconnu.

Le troisième mécanisme est celui des *collocations*, souvent héritées du patrimoine littéraire et religieux, dont la première moitié appelle la seconde de manière quasi systématique. Ainsi, الحمد لله *al-ḥamdu* appelle لله *llāhi* (« louange à Dieu »). De même, le proverbe مَنْ جَدَّ *man jadda* appelle وَجَدَّ *wajada* (« qui persévère trouve »), et لا قوة إلا بالله *lā ḥawla* projette بالله *llāhi*.

quwwata illā billāhi (« il n'y a de force qu'en Dieu »). Le proverbe arabe, comme l'invocation rituelle, est précisément l'une des formes où la prédictibilité sémantique de l'arabe atteint son intensité maximale.

Les exemples suivants illustrent *différentes configurations de prédiction sémantique* en arabe standard :

Exemple 1 — Champ sémantique racinaire : Dans un contexte éducatif, la production du mot *دَرَسَ darasa* (« il a étudié »), construit sur la racine *درس (d-r-s)*, active un réseau clos : *دَرْس dars* (« leçon »), *مَدْرَسَة madrasa* (« école »), *مُدَرِّس mudarris* (« enseignant »), *دِرَاسَة dirāsa* (« étude »). D'autres continuations, lexicalement possibles, sont statistiquement improbables dans ce contexte.

Exemple 2 — Contrainte du schème dérivationnel : Le schème *مُفْتَعَل mufta'al*, participe passif de la huitième forme, sélectionne un sens de résultat ou d'aboutissement : *مُنْتَظَر muntaẓar* (« attendu »), *مُجْتَمَع mujtama'* (« société/réunion »), *مُعْتَقَل mu'taqal* (« camp de détention »). Le schème fonctionne comme un filtre prédictif qui contraint l'interprétation catégorielle.

Exemple 3 — Polarité et attente lexicale : L'énoncé *لَيْسَ لَيْسًا فَفَقَطُ دَكِيًّا بَلْ أَيْضًا* __ *laysa faqaṭ dakiyyan bal ayḍan* __ (« Il n'est pas seulement intelligent, mais aussi __ ») contraint la suite à un adjectif mélioratif : *مُجْتَهِد mujtahid* (« assidu »), *كَرِيم karīm* (« généreux »), *مُهَذَّب muhaḍḍab* (« poli »). Un adjectif dépréciatif violerait la cohérence argumentative de la construction *فَقَطُ... لَيْسَ... بَلْ...* (certes... mais...).

Exemple 4 — Prédiction par contraste concessif : Dans *أَلَا أَنَّهُ عَلَى الرَّغْمِ مِنْ صُعُوبَةِ الْإِمْتِحَانِ، إِلَّا أَنَّهُ شُ'ŪBATI L-IMTIHĀNI, ILLĀ ANNAHU* __ (« Malgré la difficulté de l'examen, il __ »), la structure concessive ou de réussite : *نَجَحَ najaha* (« a réussi »), *تَفَوَّقَ tafawwqa* (« a excellé »).

Exemple 5 — Formules figées : L'amorce expressive *إِنَّا لِلَّهِ innā lillāhi* appelle exclusivement *وَأِنَّا إِلَيْهِ رَاجِعُونَ wa-innā ilayhi rāji'ūn* (« nous appartenons à Dieu et c'est à Lui que nous retournons »), formule prononcée à l'annonce d'un décès ou d'une nouvelle funeste. De même, *مَا شَاءَ mā shā'a* projette systématiquement *اللَّهُ Allāhu* (« ce que Dieu a voulu »), expression d'admiration et de protection contre le mauvais œil. Enfin, *إِنْ شَاءَ اللَّهُ inshā'a Allāh* (inchallah, « si Dieu le veut »), même s'il est sémantiquement prédictif, laisse ouverte la possibilité d'occurrence ou de réalisation dans le futur.

La prédiction syntaxique en arabe

La prédiction syntaxique porte sur l'anticipation du syntagme ou de la phrase qui prolonge un énoncé ou un discours en cours. Elle s'appuie, en arabe, sur un ensemble de particules de subordination, de flexions casuelles et de patrons verbaux dont la valeur prédictive est élevée. La richesse morphosyntaxique de l'arabe offre des indices de prédiction d'une précision supérieure à celle observée dans les langues à morphologie réduite.

Le premier ressort de la *prédiction syntaxique* réside dans les *particules de subordination et de modalisation*, qui modifient le mode du verbe et contraignent la structure de la proposition qui suit. Ainsi par exemple, la particule *أَنَّ an* impose le subjonctif (منصوب) au verbe qui la suit : *أُرِيدُ أَنْ أَتَعَلَّمَ urīdu an ata'allama* (« je veux apprendre »). La particule *لَنْ lan* impose également le subjonctif mais avec une valeur de négation future définitive : *لَنْ أَنْسَى lan ansā* (« je n'oublierai jamais »). La particule *لَمْ lam* impose l'apocopé (مجزوم) et projette une négation passée : *لَمْ يَخْضُرْ lam yahḍur* (« il n'est pas venu »). Dès la production de l'une de ces particules, le locuteur projette non seulement un mode verbal spécifique, mais aussi un type de proposition clairement contraint.

Le deuxième ressort tient à la *structure de la phrase nominale et verbale*. En effet, l'arabe oppose classiquement la phrase verbale (جملة فعلية, *jumla fi'liyya*), qui commence par un verbe, à la phrase nominale (جملة اسمية, *jumla ismiyya*), qui commence par un nom. Dans la phrase verbale, l'ordre canonique Verbe-Sujet-Objet (فاعل - مفعول به) est hautement prédictif : après un verbe transitif, le système projette un sujet au nominatif suivi d'un complément à l'accusatif. Dans la phrase nominale, l'apparition d'un *mubtada'* (مبتدأ, « thème ») projette un *ḥabar* (خبر, « prédicat ») qui s'accorde en genre et en nombre. Cette opposition structurelle fait du premier élément de la phrase un opérateur prédictif puissant.

Le troisième ressort est celui des *particules introductrices* (حروف, *hurūf*) qui bornent les possibilités syntaxiques et sémantiques en aval. La particule *إِنَّ inna* (« certes, en effet ») impose l'accusatif au nom qui la suit et projette un prédicat au nominatif : *إِنَّ الْعِلْمَ نُورٌ inna l-'ilma nūrun* (« certes, le savoir est une lumière »). La particule *كَانَ kāna* (« était ») impose le nominatif à son sujet et l'accusatif à son attribut : *كَانَ الطَّقْسُ جَمِيلًا kāna ṭ-ṭaqsu jamīlan* (« le temps était beau »). Chaque particule sélectionne une classe de continuations probables et en écarte d'autres.

Les exemples suivants illustrent *différents régimes de prédiction syntaxique* :

Exemple 1 — Conditionnel : Dans *إِذَا اجْتَهَدْتَ* __ *idā jtahadta*, __ (« Si tu fais des efforts, __ »), la structure projette une apodose introduite par *فَ* *fa-* : *فَسَتَنْجَحُ* *fa-satanjahu* (« alors tu réussiras »). L'omission du *fa-* après *إِذَا* dans la réponse est exceptionnelle.

Exemple 2 — Annexion (*idāfa*) : Le syntagme *كِتَابٌ* __ *kitābu* __ (« le livre de __ »), avec un nom au nominatif indéfini en position d'annexé, projette obligatoirement un nom au génitif (cas indirect) : *كِتَابُ التَّارِيخِ* *kitābu t-tārīhi* (« le livre d'histoire »), *كِتَابُ الْمُعَلِّمِ* *kitābu l-mu'allimi* (« le livre de l'enseignant »). La construction d'annexion (*إِضَافَةٌ*, *idāfa*) est l'un des mécanismes prédictifs les plus contraignants de l'arabe.

Exemple 3 — Verbe et complément d'objet : Dans *قَرَأَ* *qara'a* *الطَّالِبُ* __ *ṭ-ṭālibu* __ (« l'étudiant a lu __ »), le verbe transitif suivi du sujet projette un complément d'objet à l'accusatif (cas direct) : *الْكِتَابَ* *al-kitāba* (« le livre »), *الْمَقَالَةَ* *al-maqālata* (« l'article »), *الرِّسَالَةَ* *ar-risālata* (« la lettre »).

Exemple 4 — Relative (*al-ẓayy*) : L'énoncé *الرَّجُلُ الَّذِي* __ *ar-rajulu llaḏī* __ (« l'homme qui __ ») projette une proposition relative contenant un verbe conjugué et, éventuellement, un pronom de rappel (ضمير عائد) : *الرَّجُلُ الَّذِي قَابَلْتُهُ* *ar-rajulu llaḏī qābaltuhu* (« l'homme que j'ai rencontré »).

Exemple 5 — Construction exceptive : L'énoncé *حَضَرَ* *ḥaḍara* *إِلَّا* __ *illā* __ *الطَّالِبُ* __ *ṭ-ṭālibu* __ (« les étudiants sont venus sauf __ ») projette un nom à l'accusatif : *طَالِبًا* *ṭāliban wāḥidan* (« un seul étudiant »). La particule d'exception *إِلَّا* contraint le cas du nom excepté.

Exemple 6 — Inna et saœurs : L'amorce *لَعَلَّ* __ *la'allā* __ (« peut-être que __ ») impose l'accusatif au sujet qui suit et projette un prédicat au nominatif : *لَعَلَّ الْجَوَّ* *la'allā l-jawwa mu'tadilon* (« peut-être le temps est-il clément »).

Exemple 7 — Coordination et attente : L'énoncé *سَافَرَ* *sāfara* *إِلَى* *ilā* *لُنْدُنَا* *lunduna* *ثُمَّ* __ *thumma* __ (« il a voyagé à Londres puis __ »), l'adverbe de succession temporelle *thumma* __ projette un second événement dans la même ligne narrative : *عَادَ* *āda* *إِلَى* *ilā* *بَلَدِهِ* *baladihi* (« est retourné dans son pays »), *انْتَقَلَ* *intaqala* *إِلَى* *ilā* *بَارِيسَ* *bārīsa* (« s'est rendu à Paris »).

Exemple 8 — Négation et portée : L'amorce *مَا كَانْ* __ *mā kāna* __ (« il n'était pas ») impose un sujet au

nominatif et projette un attribut à l'accusatif : *مَا كَانْ* *mā kāna* *لِأَمْرٍ* *l-amru* *سَهْلًا* *sahlan* (« la chose n'était pas facile »). La négation de *kāna* *كان* ne modifie pas la répartition casuelle mais contraint la polarité sémantique du prédicat.

La prédiction pragmatique en arabe

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 mot 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 et culturelle. L'arabe y déploie une densité de scripts religieux, rhétoriques et sociaux qui rend la prédiction pragmatique particulièrement structurée.

Le premier axe de prédiction pragmatique tient aux *scripts rituels de salutation et d'interaction*. L'échange *السَّلَامُ عَلَيْكُمْ* *as-salāmu 'alaykum* (« que la paix soit sur vous ») appelle obligatoirement *وَعَلَيْكُمْ السَّلَامُ* *wa-'alaykumu s-salām* (« et que la paix soit sur vous »), souvent augmenté de *وَرَحْمَةُ اللَّهِ وَبَرَكَاتُهُ* *wa-rahmatu llāhi wa-barakātuhu* (« et la miséricorde et les bénédictions de Dieu »). La réponse est non seulement attendue mais religieusement prescrite, ce qui porte sa prédictibilité à un degré maximal dans de nombreux contextes sociaux. De même, après un éternuement, *الْحَمْدُ لِلَّهِ* *al-ḥamdu lillāhi* (« louange à Dieu ») appelle *يَرْحَمُكَ اللَّهُ* *yarḥamuka llāhu* (« que Dieu te fasse miséricorde »), puis *يَهْدِيكَ اللَّهُ* *yahdīkumu llāhu* (« que Dieu vous guide »).

Le deuxième axe est celui de *l'hospitalité et de la générosité rituelle*. À l'arrivée d'un invité, l'hôte prononce *أَهْلًا وَسَهْلًا* *ahlan wa-sahlan* (« soyez le bienvenu »), qui appelle de la part de l'invité *يَحْفَظُكَ اللَّهُ* *Allāhu yahfazaka* (« que Dieu vous protège ») ou *اللَّهُ يُبَارِكُ فِيكَ* *Allāhu yubāriku fika* (« que Dieu vous bénisse »). Lorsque l'hôte propose un repas, la séquence attendue est un premier refus poli de la part de l'invité — *شُكْرًا، لَا* *shukran, lā* (« merci, je ne voudrais pas vous importuner »), suivi de l'insistance de l'hôte et d'une acceptation finale. Ce schéma *d'insistance-refus-acceptation* est hautement prédictible et joue un rôle structurant dans l'interaction sociale en arabe.

Le troisième axe concerne les *formules de félicitation et de condoléances*. À l'annonce d'une nouvelle heureuse (réussite, mariage, etc.), *مَبْرُوكٌ* *mabrūk* (« félicitations ») appelle *اللَّهُ يُبَارِكُ فِيكَ* *Allāhu yubāriku fika* (« que Dieu vous bénisse »). À l'inverse, l'annonce d'un décès, *عَظَّمَ اللَّهُ أَجْرَكَ* *'azzama llāhu ajraka* (« que

Dieu augmente votre rétribution ») appelle جَزَاكَ اللهُ خَيْرًا *jazāka llāhu ḥayran* (« que Dieu vous récompense par le bien »). Ces paires formulaires sont si stabilisées que toute déviation constituerait une rupture pragmatique remarquée.

Le quatrième axe porte sur les *requêtes indirectes et la politesse*. En effet, l'arabe tend à privilégier des formulations atténuées dans les interactions formelles. Une demande directe أَعْطِنِي مَاءً *a'ṭini mā'an* (« donne-moi de l'eau ») serait perçue comme brusque dans de nombreux contextes ; la forme attendue est لَوْ سَمَحْتَ، هَلْ يُمَكِّنُكَ أَنْ تُعْطِنِي مَاءً؟ *law samaḥta, hal yumkinuka an tu'ṭiyāni mā'an?* (« s'il te plaît, pourrais-tu me donner de l'eau ? »). La présence de لَوْ سَمَحْتَ (« si tu le permets ») et du couple هَلْ يُمَكِّنُكَ (« te serait-il possible ») constitue un marqueur prédictif stable de la requête polie.

Le cinquième axe est celui des *schémas narratifs et rhétoriques*. La tradition rhétorique arabe (*balāgha*) structure le discours selon des figures hautement prédictibles. Le *saj'* (سجع, prose assonancée) impose des clausules rimant entre elles, dont la terminaison de la première prédit celle de la seconde. Les proverbes arabes fonctionnent de la même manière : مَنْ طَلَبَ الْعُلَا *man ṭalaba l-'ulā* (« qui recherche l'élévation ») appelle سَهَرَ اللَّيَالِي *sahira l-layālī* (« veille les nuits »). De même, الصَّبْرُ مِفْتَاحُ *aṣ-ṣabru miftāḥu* (« la patience est la clé de... ») projette systématiquement الْفَرْجِ *al-farajī* (« du soulagement »).

Le sixième axe concerne les *particules discursives* qui orientent la suite attendue de la conversation. Ainsi, إِنَّ شَاءَ اللهُ *in shā'a llāhu* (« si Dieu le veut ») ne fonctionne pas comme une simple formule pieuse : c'est un marqueur de modalité épistémique qui signale l'intention du locuteur tout en atténuant l'engagement, projetant une réponse de type اللهُ يُسَوِّلُ *Allāhu yusahhilu* (« que Dieu facilite ») ou بِإِذْنِ اللهِ *bi-īdni llāhi* (« avec la permission de Dieu »). De même, la particule اللهُ أَعْلَمُ *Allāhu a'lam* (« Dieu sait mieux ») signale la clôture d'une opinion personnelle et projette soit l'acquiescement, soit un débat nuancé.

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

Exemple 1 — Rencontre entre collègues : كَيْفَ حَالُكَ؟ *kayfa ḥāluka?* (« comment vas-tu ? ») projette بِحَمْدِ اللهِ، بِخَيْرٍ *al-ḥamdu lillāhi, bi-ḥayr* (« grâce à Dieu, bien »), suivi de وَأَنْتَ؟ *wa-anta?* (« et toi ? »). La mention de Dieu dans la réponse est culturellement attendue et constitue un marqueur pragmatique stable.

Exemple 2 — Invitation à un repas : تَقْضَلُنَا عَلَى الْعِشَاءِ *tafaddal 'indanā 'alā l-'ashā'i* (« faites-nous l'honneur

de dîner chez nous ») appelle d'abord لا اللهُ يُبَارِكُ فِيكَ، أَرِيدُ إِتْفَاكَ *Allāhu yubāriku fika, lā urīdu ithqālaka* (« que Dieu vous bénisse, je ne voudrais pas vous déranger »), avant acceptation après insistance.

Exemple 3 — Réception d'un cadeau : Le don, accompagné de هَذِهِ هَدِيَّةٌ بَسِيطَةٌ *hāḍihi hadiyyatun basiṭatun* (« c'est un modeste cadeau »), appelle مَا كَانَ يَلْزَمُ *mā kāna yalzamu* (« il ne fallait pas ») avant l'acceptation, souvent assortie de جَزَاكَ اللهُ خَيْرًا *jazāka llāhu ḥayran* (« que Dieu vous récompense par le bien »).

Exemple 4 — Excuse et réparation : تَأَخَّرْتُ، عُذْرًا *'udran, ta'aḥḥartu* (« pardon, je suis en retard ») projette لَا أَنَا أَيْضًا وَصَلْتُ تَوًّا *lā ba'sa, anā ayḍan waṣaltu tawwan* (« pas de problème, moi aussi je viens juste d'arriver »), même si l'interlocuteur attend depuis longtemps.

Exemple 5 — Clôture téléphonique : La séquence طَيِّبَ، إِلَى الْقَاءِ *ṭayyib, ilā l-liqā'i* (« bien, à la prochaine ») projette مَعَ السَّلَامَةِ *ma'a s-salāma* (« avec la paix/au revoir »), suivi d'الله يُسَلِّمُكَ *Allāhu yusallimuka* (« que Dieu te préserve »). Interrompre l'appel sans cette séquence serait jugé impoli.

Exemple 6 — Négociation au marché : L'annonce du prix par le vendeur بِعِشْرِينَ دِينَارًا *bi-'ishrīna dīnāran* (« vingt dinars ») projette هَلْ يُمَكِّنُ تَخْفِيفًا؟ *ghālin ghālin* (« c'est trop cher, une réduction est-elle possible ? »), suivi d'un marchandage en plusieurs tours. L'absence de négociation serait, dans ce contexte, culturellement marquée.

Exemple 7 — Demande de service : L'amorce عِنْدِي طَلْبٌ *'indī ṭalabun basiṭun, in lam yakun fihī iz'ājūn* (« j'ai une petite demande, si cela ne vous dérange pas ») annonce une requête dont le contenu sera différé, précédé d'une justification contextuelle. La structure est prédictible dans sa forme 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 de l'arabe 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 de l'arabe. En effet, la grammaire prédictive de l'arabe, 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 de l'arabe. Cette convergence entre linguistique prédictive 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.

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 l'arabe comme une langue parmi d'autres, sans prise en compte systématique des trois traits structurels que nous avons identifiés : la morphologie dérivationnelle, la flexion casuelle et la contextualité culturelle.

Cette homogénéisation génère des modèles qui, bien qu'opérationnels, reflètent mal les profils prédictifs propres à l'arabe et produisent des énoncés formellement corrects mais pragmatiquement décalés.

La première contribution d'une *grammaire prédictive de l'arabe* à la conception de tels modèles concerne la *tokenisation*. Une segmentation aveugle en sous-mots, héritée du traitement de l'anglais, fragmente arbitrairement les formes dérivées et brise la solidarité racine-schème qui structure le mot arabe. Une *tokenisation informée par la linguistique prédictive* intégrerait, à l'inverse, la dimension de la racine trilitère, la cohésion des schèmes dérivationnels et la typicité des formules figées. Elle reconnaîtrait que la probabilité de la séquence *maktaba* مكتبة n'est pas réductible au produit des probabilités isolées de ses composants graphiques, mais qu'elle exprime une *unité lexicale structurée* dont

la cohésion morphosémantique est elle-même porteuse d'information.

La deuxième contribution concerne le *traitement des particules et des flexions casuelles*. Les particules de subordination (أَنَّ, لَنْ, لَمْ, إِنَّ, كَأَنَّ, لَكِنَّ, لَعَلَّ, لَيْتَ) et les désinences casuelles (*fatha, damma, kasra*) sont des opérateurs prédictifs à longue portée, dont la fonction dépasse la simple marque syntaxique. Un *modèle prédictif adapté à l'arabe* 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 structurels, 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 à la salutation السلام عليكم, un simple مَرْحَبًا peut être techniquement correct mais culturellement inadéquat. Il manque au locuteur arabe l'attente de la réponse rituelle وعليكم السلام ورحمة الله. De même, un modèle qui accepte une invitation dès la première offre, sans passer par la séquence conventionnelle de refus poli, rompt un script social stabilisé.

Ces décalages, parfois imperceptibles aux évaluations quantitatives, grèvent lourdement l'acceptabilité sociale des productions en contexte arabophone. Une *grammaire prédictive pragmatique* devrait fournir au concepteur de modèle de langage 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, les formules de piété, le jeu des rapports entre *'adab* (أدب, « littérature, éducation morale ») et sacralité, la valorisation de la patience (صبر), la préférence pour l'indirection et l'atténuation, autant de traits qui ne sont pas des ornements pragmatiques surajoutés à une grammaire neutre, mais des

composantes intrinsèques de la prédictibilité de l'arabe. 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 signalé 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 (racines, schèmes, particules, flexions, 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.

L'arabe, par la rigueur de sa morphologie 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 de l'arabe peut, *mutatis mutandis*, être appliqué à toute langue dont la richesse est aujourd'hui insuffisamment captée par les modèles de langage dominants.

Conclusion

La linguistique prédictive apparaît non seulement comme un outil d'analyse, mais aussi comme un principe organisateur permettant de reconfigurer en profondeur la compréhension de la grammaire arabe. Les approches descriptives traditionnelles reposaient sur une logique classificatoire, fondée sur l'énumération de règles et la stabilisation de paradigmes. La perspective prédictive introduit un déplacement décisif en orientant l'analyse vers les dynamiques d'anticipation qui structurent l'activité langagière. Elle met en lumière ce qui rend certaines

suites plus plausibles que d'autres, ce qui façonne les attentes conjointes du locuteur et de l'interlocuteur, et ce qui confère à la langue son caractère ordonné malgré l'apparente infinité de ses réalisations possibles.

Dans ce cadre, la langue n'est plus envisagée comme un répertoire ouvert de combinaisons équivalentes, mais comme un espace stratifié où les occurrences sont distribuées selon des degrés de probabilité dépendant du contexte. La linguistique prédictive ne remplace pas les modèles théoriques antérieurs, mais en propose une reformulation plus intégrative. Elle leur donne un fondement transversal (la prédiction), capable de rendre compte à la fois des régularités les plus stables et des variations les plus subtiles observables dans les usages effectifs.

L'organisation en trois niveaux d'analyse, sémantique, syntaxique et pragmatique, ne relève pas d'un simple choix méthodologique destiné à clarifier l'exposé. Elle correspond à la structuration interne de la compétence linguistique en arabe, telle qu'elle se manifeste dans les processus d'anticipation. Ces niveaux ne fonctionnent pas de manière isolée, mais s'entrelacent dans une dynamique unifiée où chaque indice, qu'il soit lexical, grammatical ou discursif, contribue à orienter la construction du sens et la projection des formes à venir.

Au niveau sémantique, nous avons vu la prédiction opérer à l'échelle la plus fine, celle de la racine et du mot, à travers la productivité du système racinaire, la régularité des schémas dérivationnels, la cristallisation des formules et la contrainte exercée par les solidarités lexicales.

Au niveau syntaxique, la prédiction s'est déployée à l'échelle du syntagme et de la phrase, mobilisant les particules de subordination modale (أَنَّ, لَنْ, لَمْ), les particules introductrices (إِنَّ et ses sœurs), la construction d'annexion (إِضَافَةٌ), l'opposition entre phrase verbale et phrase nominale, ainsi que la flexion casuelle comme marqueur de fonction prédictif.

Au niveau pragmatique enfin, la prédiction a pris la forme d'une anticipation de séquences rituelles, d'actes de langage attendus et de schémas conversationnels stabilisés, où se manifestent les formules religieuses, les stratégies d'hospitalité, les codes de politesse indirecte et les figures rhétoriques héritées de la tradition littéraire arabe.

Ces trois niveaux, loin d'être étanches, s'entredéterminent en permanence : le choix d'une particule de subordination suppose une anticipation

syntaxique, de même que la reconnaissance d'une racine oriente une prédiction sémantique, et l'identification d'une formule rituelle d'ouverture projette un schéma pragmatique complet et une structuration particulière de l'interaction.

Il convient de souligner que les ressorts prédictifs mis en évidence forment le cœur opératoire de la langue et le socle de son intelligibilité. Là où une grammaire descriptive pourrait traiter les particules de subordination 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 propositions. 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 de l'arabe qui ignorerait la convention de la réponse à la salutation 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 de l'arabe qui ne soit, indissociablement, maîtrise de la culture.

La grammaire prédictive que nous avons esquissée s'adresse à un double public. Pour l'apprenant de l'arabe, elle offre une méthode d'appropriation progressive et contextualisée, qui rompt avec la traditionnelle accumulation de règles et de paradigmes. 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.

Elle rejoint également les pratiques traditionnelles les plus éprouvées de l'enseignement de l'arabe, qui ont toujours valorisé la mémorisation de versets coraniques, de poèmes préislamiques 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.

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 arabophone 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. 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 de l'arabe.

En ce sens, la grammaire prédictive de l'arabe 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. 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.

La langue arabe, par le défi qu'elle lance aux paradigmes dominants, n'est pas seulement un objet d'étude parmi d'autres : elle est, du point de vue de la linguistique prédictive, un laboratoire privilégié où se joue la possibilité même d'une intelligence artificielle véritablement plurilingue et interculturelle.

Références

- Abdelhady, N., Elseman, I. E., & Soliman, T. H. A. (2024). A real-time predicting online tool for detection of people's emotions from Arabic tweets based on big data platforms. *Journal of Big Data*, 11(1), 171.
- Aboutaib, A., El Allaoui, A., Zeroual, I., & Dadi, E. W. (2023, May). Punctuation Prediction for the Arabic language. In *Proceedings of the 6th International Conference on Networking, Intelligent Systems & Security* (pp. 1-4).
- Al Wazrah, A. A., Altamimi, A., Aljasim, H., Alshammari, W., Al-Matham, R., Elnashar, O., ... & AlOsaimy, A. (2025, January). Evaluation of large language models on arabic punctuation prediction. In *Proceedings of the 1st Workshop on NLP for Languages Using Arabic Script* (pp. 144-154).
- Alabid, N. N. (2026). A Hybrid Context-Knowledge Representation Model for Arabic Next Word Prediction. *Journal of Al-Qadisiyah for Computer Science and Mathematics*, 18(1), 114-123.
- Mokhamed, T., Harous, S., Hussein, N., & Ismail, H. (2024). Comparative analysis of Deep Learning and Machine Learning algorithms for emoji prediction from Arabic text. *Social Network Analysis and Mining*, 14(1), 67.
- Aladeemy, A. A., Aldhyani, T. H., Alzahrani, A., Alzahrani, E. M., Khalaf, O. I., Alsubari, S. N., ... & Al-Adhaileh, M. H. (2024). Machine learning algorithms for predicting and analyzing arabic sentiment. *SN Computer Science*, 5(8), 1132.
- Alghazo, S., & Jarrah, M. (2023). No adjective ordering preferences in Jordanian Arabic grammar. *Journal of Psycholinguistic Research*, 52(5), 1633-1667.
- Al-Jarf, R. (2025). DeepSeek, Google translate and Copilot's translation of Arabic grammatical terms used metaphorically. *Journal of Computer Science and Technology Studies*, 7(3), 46-57.
- Aljassmi, M. A., Warrington, K. L., McGowan, V. A., White, S. J., & Paterson, K. B. (2022). Effects of word predictability on eye movements during Arabic reading. *Attention, Perception, & Psychophysics*, 84(1), 10-24.
- Alrehili, A., & Alhothali, A. (2025). Towards the Development of Balanced Synthetic Data for Correcting Grammatical Errors in Arabic: An Approach Based on Error Tagging Model and Synthetic Data Generating Model. *arXiv preprint arXiv:2502.05312*.
- Alshdaifat, E., Al-shdaifat, A., & Alsarhan, A. (2024). Machine Learning Algorithms for Attitude Prediction from Arabic Text: Detecting Student Attitude towards Online Learning. *International Journal of Interactive Mobile Technologies*, 18(12).
- Alzahrani, A. (2025). The predictive processing of number information in L1 and L2 Arabic speakers. *Languages*, 10(2), 25.
- Ambreen, S., Iqbal, M., Asghar, M. Z., Mazhar, T., Khattak, U. F., Khan, M. A., & Hamam, H. (2024). Predicting customer sentiment: the fusion of deep learning and a fuzzy system for sentiment analysis of Arabic text. *Social Network Analysis and Mining*, 14(1), 206.
- Asli-Badarneh, A., Hipfner-Boucher, K., Bumgardner, X. C., Aljanaideh, R., & Saiegh Haddad, E. (2023). Narrative microstructure and macrostructure skills in Arabic diglossia: The case of Arab immigrant children in Canada. *International Journal of Bilingualism*, 27(3), 349-373.
- Benlaaraj, O., El Jaafari, I., Ellahyani, A., & Boutaayamou, I. (2022, October). Prediction of suicidal ideation in a new arabic annotated dataset. In *2022 9th International Conference on Wireless Networks and Mobile Communications (WINCOM)* (pp. 1-5). IEEE.
- Bensalem, R., Haddar, K., & Blache, P. (2023). An Arabic probabilistic parser based on a property grammar. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 22(10), 1-25.
- Bozza, C. (2024). On the grammaticalization of ġādi in Moroccan Arabic: new insights. *Journal of African Languages and Linguistics*, 45(1), 1-43.
- Ouali, H. (2022). The imperfective verb and the progressive aspect in Arabic. *Perspectives on Arabic linguistics XXXIII*, 121-146.
- Chamsine, C. (2018). Traduire les émotions. Paris : Éditions L'Harmattan. Collection Traductologie.
- Essa, N., Elgayar, M., & El-Daydamony, E. (2025). Arabic Grammar Correction for Arabic Text Summaries. *Mansoura Journal for Computer and Information Sciences*, 20(2), 1-16.
- 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.

Hamed, I., El Bolock, A., Herbert, C., Abdennadher, S., & Vu, N. T. (2022). The Who in Code-Switching: A Case Study for Predicting Egyptian Arabic-English Code-Switching Levels Based on Character Profiles. *International Journal of Asian Language Processing*, 32(01), 2250010.

Layes, S., & Bouakkaz, T. (2022). Predicting word and pseudoword reading in Arabic-speaking children: the independent contributions of phonological and morphological awareness and visual attention. *British Journal of Special Education*, 49(2), 230-260.

López-Riobóo, E., & Martínez-Castilla, P. (2026). Exploring the Predictive Role of Lexical Stress Discrimination in the Phonological and Grammatical Skills of Teenagers With Down Syndrome. *Journal of Intellectual Disability Research*, 70(1), 48-57.

Mansour, Y., Elnagar, A., & Yagi, S. (2023, March). Punctuation prediction for the Arabic language. In *Doctoral Symposium on Computational Intelligence* (pp. 579-592). Singapore: Springer Nature Singapore.

Matar, S., & Marantz, A. (2025). Neural bases of proactive and predictive processing of meaningful subword units in speech comprehension. *Journal of Neuroscience*, 45(7).

Mukmin, M., Hidayah, N., Haya, A. S., Ali, N. J. M., & Abdullah, A. W. (2025). Determining Arabic language proficiency: An examination of grammar mastery, active usage, and discipline of linguistic. *Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab*, 17(1), 178-194.

Raheem, H. A. (2026). An LSTM-based approach for predicting the next word in Arabic language. *Journal of Al-Qadisiyah for Computer Science and Mathematics*, 18(1), 270-280.

Sabaneh, K., Salameh, M. A., Khaleel, F., Herzallah, M. M., Natsheh, J. Y., & Maree, M. (2023, November). Early risk prediction of depression based on social media posts in Arabic. In *2023 IEEE 35th International Conference on Tools with Artificial Intelligence (ICTAI)* (pp. 595-602). IEEE.

Taha-Thomure, H., & Harb, M. (2022). Arabic grammar at the advanced level: A paradigm shift. In *Teaching and Learning Arabic Grammar* (pp. 232-252). Routledge.

Zolkapli, R. B. M., Kenali, S. F. M., Hadi, N. F. A., Jaafar, A. J., Rahim, M. I., Ahmad, N. H. F., ... & Khairuddin, N. A. (2026). Contrastive Analysis between English and Arabic Grammar. *Malaysian Journal of Social Sciences and Humanities (MJSSH)*, 11(2), e003834-e003834.