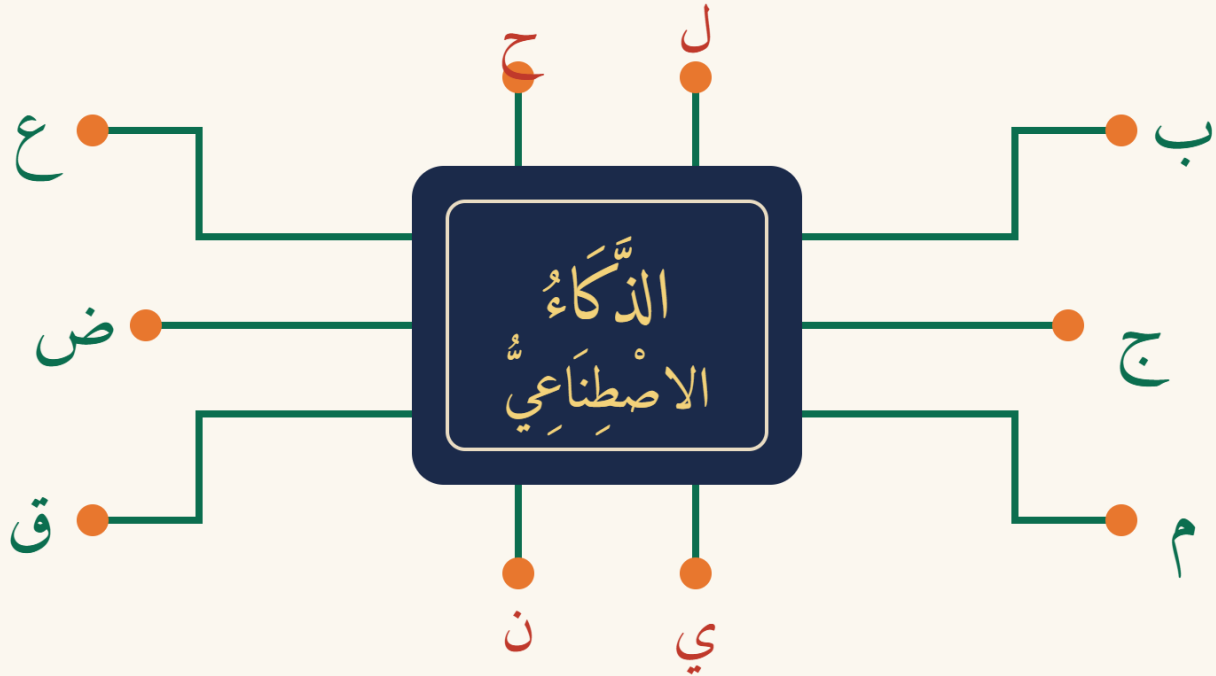


# L'arabe et l'intelligence artificielle

Cours d'arabe en seconde



La machine apprend la langue à partir des données

*L'intelligence artificielle apprend aussi l'arabe, lettre après lettre, donnée après donnée.*

## Découvrir le document : IA arabe 2nde

Voici un court article écrit pour ce cours. Comme il s'agit du document de la leçon, il est **entièrement vocalisé**. Dans la fiche d'exercices, les textes d'entraînement seront **peu ou pas vocalisés**, comme dans la presse arabe : c'est l'habitude à prendre au lycée. Le mot « علم » reste volontairement sans voyelles, puisque c'est justement l'exemple de l'ambiguïté.

### حِينَ تَتَعَلَّمُ الآلَةُ الْعَرَبِيَّةُ

نَسْتَعْمِلُ الْيَوْمَ تَطْبِيقَاتِ الذِّكَاةِ الْاصْطِنَاعِيِّ كُلِّ يَوْمٍ: نُتَرْجِمُ بِهَا نُصُوصًا، وَنَطْرَحُ عَلَيْهَا أَسْئَلَةً، وَنَطْلُبُ مِنْهَا أَنْ تَكْتُبَ رِسَالَةً. وَلَكِنْ كَيْفَ «تَفْهَمُ» هَذِهِ الْبَرَامِجُ اللُّغَةَ الْعَرَبِيَّةَ؟ لَا تَفْهَمُ الآلَةُ كَمَا يَفْهَمُ الْإِنْسَانُ، بَلْ تَتَعَلَّمُ مِنْ كَمِّيَّاتٍ كَبِيرَةٍ مِنَ الْبَيِّنَاتِ، أَيَّ مِنْ مَلَائِينَ الْجُمْلِ الْمَكْتُوبَةِ عَلَى الْإِنْتَرْنِتِ. وَهُنَا تَبْدَأُ التَّحْدِيَّاتُ. أَوَّلًا، تُكْتُبُ الْعَرَبِيَّةُ عَادَةً بِلاَ تَشْكِيلٍ، فَكَلِمَةُ «علم» مَثَلًا قَدْ تُقْرَأُ «عِلْمٌ» أَوْ «عِلْمٌ» أَوْ «عِلْمٌ»، وَعَلَى الْآلَةِ أَنْ تَخْتَارَ الْقِرَاءَةَ

الصَّحِيحَةَ مِنَ السِّيَاقِ. ثَانِيًا، يَتَكَلَّمُ النَّاسُ لَهَجَاتٍ كَثِيرَةً تَخْتَلِفُ مِنَ الْمَغْرِبِ إِلَى الْعِرَاقِ، وَيَكْتُبُونَ بِهَا عَلَى شَبَكَاتِ التَّوَاصُلِ الْاجْتِمَاعِيِّ. ثَالِثًا، النُّصُوصُ الْعَرَبِيَّةُ الْمُتَاحَةُ عَلَى الْإِنْتَرْنِتِ أَقَلُّ عَدَدًا مِنَ النُّصُوصِ الْإِنْجِلِيزِيَّةِ، فَتَتَعَلَّمُ الْآلَةُ مِنْهَا أَقَلَّ.

لِلتَّرْجَمَةِ الْآلِيَّةِ مَزَايَا وَاضِحَةٌ: فَهِيَ سَرِيعَةٌ وَمَجَانِيَّةٌ فِي أَغْلَبِ الْأَحْيَانِ، وَتُسَاعِدُ عَلَى فَهْمِ الْفِكْرَةِ الْعَامَّةِ لِنَصٍّ. لَكِنَّهَا تُخْطِئُ أحيانًا فِي الْمَعْنَى، وَقَدْ تَحْمِلُ تَحِيْزَاتٍ: فَإِذَا كَانَتْ الْبَيِّنَاتُ تُرْبِطُ مِهْنَةً مَا بِالرِّجَالِ أَكْثَرَ مِنَ النِّسَاءِ، قَدْ تَكْتُبُ الْآلَةُ «طَبِيب» حَتَّى عِنْدَمَا يَكُونُ الْحَدِيثُ عَنْ امْرَأَةٍ. وَيُمْكِنُ لِبَعْضِ الْبَرَامِجِ الْيَوْمَ أَنْ تُنْتِجَ صُورًا وَأَصْوَاتًا مُزَيَّفَةً تُشَبِّهُ الْحَقِيقَةَ.

لِذَلِكَ يَحْتَاجُ الْمُسْتَخْدِمُ إِلَى التَّفَكِيرِ النَّقْدِيِّ: أَنْ يَتَحَقَّقَ مِنَ الْمَصَادِرِ، وَأَنْ يُقَارِنَ بَيْنَ أَكْثَرِ مِنْ نَصٍّ، وَأَنْ يَتَذَكَّرَ أَنَّ الْآلَةَ قَدْ تَكْتُبُ جُمْلًا جَمِيلَةً وَلَكِنْ خَاطِئَةً.

Traduction : Quand la machine apprend l'arabe. Aujourd'hui, nous utilisons chaque jour des applications d'intelligence artificielle : nous traduisons des textes avec elles, nous leur posons des questions et nous leur demandons d'écrire un message. Mais comment ces programmes « comprennent »-ils la langue arabe ? La machine ne comprend pas comme l'être humain : elle apprend à partir de grandes quantités de données, c'est-à-dire de millions de phrases écrites sur internet. Et c'est là que commencent les défis. Premièrement, l'arabe s'écrit d'habitude sans vocalisation : le mot « علم », par exemple, peut se lire « science », « drapeau » ou « il a su », et la machine doit choisir la bonne lecture d'après le contexte. Deuxièmement, les gens parlent de nombreux dialectes qui varient du Maroc à l'Irak, et ils les écrivent sur les réseaux sociaux. Troisièmement, les textes arabes disponibles sur internet sont moins nombreux que les textes anglais : la machine en apprend donc moins. La traduction automatique a des avantages évidents : elle est rapide, le plus souvent gratuite, et elle aide à comprendre l'idée générale d'un texte. Mais elle se trompe parfois de sens et peut véhiculer des biais : si les données associent un métier aux hommes plus qu'aux femmes, la machine risque d'écrire « médecin » au masculin même lorsqu'il s'agit d'une femme. Certains programmes peuvent aujourd'hui produire des images et des voix truquées qui ressemblent à la réalité. C'est pourquoi l'utilisateur a besoin d'esprit critique : vérifier les sources, comparer plusieurs textes et se rappeler que la machine peut écrire de belles phrases... mais fausses.

## Comprendre le document

1. Comment la machine apprend-elle une langue ? Réponse : à partir de très grandes quantités de données, c'est-à-dire de millions de phrases écrites (مِنْ كَمِّيَّاتٍ كَبِيرَةٍ مِنْ) (الْبَيِّنَاتِ).

2. Quels sont les trois défis cités ? Réponse : l'écriture sans voyelles, la diversité des dialectes et le nombre plus faible de textes arabes en ligne.
3. Citez deux avantages de la traduction automatique. Réponse : elle est rapide et souvent gratuite ; elle aide à saisir l'idée générale.
4. Quel exemple de biais le texte donne-t-il ? Réponse : un métier traduit au masculin même quand il s'agit d'une femme.
5. Quelle attitude l'auteur recommande-t-il ? Réponse : l'esprit critique, **التَّفْكِيرُ النَّقْدِيُّ** : vérifier les sources et comparer.

## Le vocabulaire de l'IA en arabe

Ces mots reviennent dans tous les articles sur le numérique. Apprenez-les avec leur pluriel quand il est irrégulier (noté « ج. » pour **جَمْعٌ**, pluriel).

| Arabe                      | Français                              |
|----------------------------|---------------------------------------|
| الذِّكَاؤُ الاصْطِنَاعِيُّ | l'intelligence artificielle           |
| رَقْمِيّ                   | numérique                             |
| تَطْبِيقٌ ج. تَطْبِيقَاتٌ  | une application                       |
| بَيِّنَاتٌ                 | des données                           |
| حَاسُوبٌ                   | un ordinateur                         |
| بَرْنَامِجٌ ج. بَرَنَامِجٌ | un programme, un logiciel             |
| خَوَازِمِيَّةٌ             | un algorithme                         |
| آلَةٌ                      | une machine                           |
| آلِيّ                      | automatique                           |
| التَّرْجَمَةُ الْآلِيَّةُ  | la traduction automatique             |
| مُسَاعِدٌ صَوْتِيّ         | un assistant vocal                    |
| مُسْتَخْدِمٌ               | un utilisateur                        |
| لَهْجَةٌ ج. لَهَجَاتٌ      | un dialecte                           |
| الْفُصْحَى                 | l'arabe standard                      |
| التَّشْكِيلُ               | la vocalisation (les voyelles brèves) |
| السِّيَاقُ                 | le contexte                           |
| نَصٌّ ج. نُصُوصٌ           | un texte                              |
| مُحْتَوًى                  | un contenu                            |

|                            |                            |
|----------------------------|----------------------------|
| تَحَدٍّ ج. تَحَدِّيَاتٌ    | un défi                    |
| مِيزَةٌ ج. مَزَايَا        | un avantage                |
| عَيْبٌ ج. عُيُوبٌ          | un inconvénient, un défaut |
| تَحِيْزٌ                   | un biais, un parti pris    |
| خَطَأٌ ج. أخطاءٌ           | une erreur                 |
| مَصْدَرٌ ج. مَصَادِرُ      | une source                 |
| مَوْثُوقٌ                  | fiable                     |
| مُزَيَّفٌ                  | truqué, falsifié           |
| تَحَقَّقَ مِنْ             | vérifier                   |
| التَّفَكُّيرُ النَّقْدِيُّ | l'esprit critique          |

### Définition

الذكاء الاصطناعي associe deux mots : ذكاء (l'intelligence, racine ذ ك و) et l'adjectif اصطناعي (artificiel), formé sur اصطناع, masdar de la forme VIII اصطنع (fabriquer), de la racine ص ن ع que l'on retrouve dans صناعة (l'industrie). On l'abrège parfois en français « IA ».

### Culture

Le mot « algorithme » vient du nom du mathématicien الخوارزمي (al-Khwārizmī), qui travaillait à Bagdad au IX<sup>e</sup> siècle. Son nom, latinisé au Moyen Âge, a donné « algorithme », et le mot « algèbre » vient de الجبر, un terme présent dans le titre de l'un de ses ouvrages. En arabe moderne, « algorithme » se dit خوارزمية : la boucle est bouclée ! Par ailleurs, l'UNESCO célèbre chaque 18 décembre la Journée mondiale de la langue arabe.

## Pourquoi l'arabe est un défi pour la machine

Le document cite trois difficultés. Elles concernent aussi les apprenants : les comprendre vous aidera à mieux lire l'arabe.

## Trois défis pour la machine

1

بِلَا تَشْكِيلٍ

علم

un mot,  
plusieurs lectures

**Pas de voyelles**

2

لَهْجَاتٌ كَثِيرَةٌ

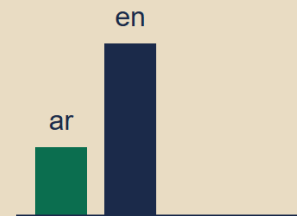


du Maghreb au Golfe

**Des dialectes variés**

3

بَيِّنَاتٌ أَقَلُّ



moins de textes

**Peu de données**

Les trois défis cités par le document.

### 1. Une écriture sans voyelles brèves

Dans la presse et sur internet, l'arabe s'écrit sans التَّشْكِيلُ. Un lecteur humain devine les voyelles grâce au السِّيَاقُ (contexte) ; la machine doit faire de même, à partir de statistiques.

### Un mot non vocalisé, trois lectures possibles

علم

عِلْمٌ

une science

عَلَمٌ

un drapeau

عَلِمَ

il a su

Un même squelette de lettres, trois mots différents.

### Exemple

رَفَعَ اللَّاعِبُونَ الْعَلَمَ بَعْدَ الْمُبَارَاةِ.

Les joueurs ont levé le drapeau après le match.

عَلِمَ الطَّالِبُ أَنَّ الْامْتِحَانَ غَدًا.

L'élève a appris (a su) que l'examen avait lieu le lendemain.

## 2. Une langue, beaucoup de dialectes

À côté de **الفُصْحَى** (l'arabe standard, celui de l'école, de la presse et de ce site), chaque région parle sa **لَهْجَة** (dialecte). Sur les réseaux sociaux, beaucoup écrivent en dialecte, parfois même en lettres latines avec des chiffres : un casse-tête pour la machine.

## 3. Moins de données

Plus une machine dispose de textes, mieux elle apprend. Or les textes arabes disponibles en ligne sont moins nombreux que les textes anglais : la qualité des outils peut donc être inégale selon les langues.

## Exprimer avantages et inconvénients avec esprit critique

Pour parler de l'IA de façon nuancée, deux outils de langue suffisent au niveau B1.

### Règle

**L'adjectif de relation (النَّسَبَةُ)** : on ajoute **يَّ** à un nom pour former un adjectif, et **يَّة** au féminin. L'adjectif s'accorde avec son nom (genre, nombre, cas, article). Un pluriel de choses s'accorde au féminin singulier.

| Nom      | Adjectif masculin | Adjectif féminin | Sens                  |
|----------|-------------------|------------------|-----------------------|
| رَقْمٌ   | رَقْمِيٌّ         | رَقْمِيَّةٌ      | chiffre → numérique   |
| آلَةٌ    | آلِيٌّ            | آلِيَّةٌ         | machine → automatique |
| صَوْتُ   | صَوْتِيٌّ         | صَوْتِيَّةٌ      | son, voix → vocal     |
| نَقْدٌ   | نَقْدِيٌّ         | نَقْدِيَّةٌ      | critique → critique   |
| مَجَّانٌ | مَجَّانِيٌّ       | مَجَّانِيَّةٌ    | gratuité → gratuit    |

### Exemple

التَّرْجَمَةُ الْآلِيَّةُ سَرِيعَةٌ.

La traduction automatique est rapide.

المصادر الموثوقة ضرورية.

Les sources fiables sont indispensables (pluriel de choses → accord au féminin singulier).

### Règle

**Peser le pour et le contre** : مِنْ مَزَايَا... أَنْ... (parmi les avantages de... il y a le fait que...), مِنْ عِيُوبٍ... أَنْ... (parmi les inconvénients...), وَمِنْ جِهَةٍ أُخْرَى... (d'un côté... de l'autre...). Après أَنْ, le sujet est au cas direct (accusatif) ou devient un pronom suffixe.

### Exemple

مِنْ مَزَايَا التَّرْجَمَةِ الْإِلَيَّةِ أَنَّهَا سَرِيعَةٌ، وَمِنْ عِيُوبِهَا أَنَّهَا تُخْطِئُ أحيانًا.

Parmi les avantages de la traduction automatique, il y a sa rapidité ; parmi ses inconvénients, le fait qu'elle se trompe parfois.

مِنْ جِهَةٍ، يُسَاعِدُنَا الذِّكَاءُ الْإِصْطِنَاعِيُّ عَلَى التَّعَلُّمِ، وَمِنْ جِهَةٍ أُخْرَى، قَدْ يُعْطِينَا مَعْلُومَاتٍ خَاطِئَةً.

D'un côté, l'intelligence artificielle nous aide à apprendre ; de l'autre, elle peut nous donner des informations fausses.

### Méthode

1. Lire le texte produit par la machine sans le recopier aussitôt.
2. Chercher la source de l'information (المصدر).
3. Comparer avec une autre source fiable (مصدر موثوق).
4. Corriger les erreurs et rédiger sa propre version.

## Les pièges à éviter

### Attention

- الإِصْطِنَاعِيُّ commence par une **alif de liaison** (forme VIII), sans hamza : n'écrivez pas الإِصْطِنَاعِي. La racine est ص ن ع, mais le ت de la forme VIII devient ط après

ص.

- الذَّكَاءُ se termine par une hamza seule sur la ligne après alif ; ne confondez pas avec l'adjectif ذَكِيٌّ (intelligent) : هَاتِفٌ ذَكِيٌّ signifie « un smartphone ».
- مَزَايَا (avantages) est invariable à la fin : on n'y ajoute pas de tanwin.
- Un pluriel de choses se reprend au féminin singulier : هَذِهِ التَّطَبِيقَاتُ مُفِيدَةٌ, et non مُفِيدُونَ.

### À retenir

- L'IA apprend l'arabe à partir de الْبَيِّنَاتُ ; ses défis : écriture sans voyelles, dialectes, données moins nombreuses.
- La traduction automatique a des مَزَايَا (rapide, gratuite) et des عُيُوبٌ (erreurs de sens, biais).
- L'adjectif de relation se forme avec يٍّ / يٍّ et s'accorde avec son nom.
- Pour nuancer : مِنْ مَزَايَا... أَنَّ, مِنْ عُيُوبٍ... أَنَّ, مِنْ جِهَةٍ... وَمِنْ جِهَةٍ أُخْرَى.
- Face à un contenu généré : التَّفَكُّيرُ النَّقْدِيُّ, vérifier et comparer les sources.

## Pour aller plus loin

Entraînez-vous avec les exercices sur l'arabe et l'intelligence artificielle, puis vérifiez vos connaissances avec le QCM du chapitre. Pour élargir le sujet, le chapitre citoyenneté et mondes virtuels prolonge cette réflexion en Terminale.