



LES PROJETS WIKIMÉDIA EN DIALECTES ARABES MAGHRÉBINS

Houcemeddine Turki^{a,*}, Mohammed Bachounda^{b,^}, Nacim Aboura^{c,^}, Emad Adel^{d,*}, Anass Sedrati^{e,+},
Ismael Zniber^{f,g,+}, Reda Benkhadra^{h,+}, Rafik Zribiⁱ, Tariq Daoudaj, Reda Kerbouche^{j,^}, Mekkia
Bennama^{k,^}, Ramzi Hachani^l

“À l’âme de mon père et à celle de mon beau-frère...”

Houcemeddine Turki



À PROPOS DE CETTE PRÉSENTATION

- Cette présentation peut être considérée comme un aperçu des discussions faites par des maghrébins ainsi que par des étrangers à propos des projets Wikimedia en arabe maghrébin pendant plus que 10 ans.
- Cette présentation est aussi issue des discussions faites pendant la WikiConvention Francophone lors de la présentation des travaux suivants:
 - Reda Benkhadra et Anass Sedrati, *Défis linguistiques des pays francophones sur Wikipédia, exemple du Maroc*
 - Lucas Lévêque, *Les langues de France sur les projets Wikimedia*
 - Xabier Cañas, *Langue basque*
 - Pierre Yves Beaudoin et Nicolas Vigneron, *Wikisource*
 - Rémy Gerbet, Lucas Lévêque et Nicolas Vion, *Lingua Libre*
 - Benoît Evellin, *Découverte de l'outil de traduction de contenu*
 - Houcemeddine Turki, *Révision d'articles de Wikipédia assistée par des spécialistes*
 - Sylvain Boissel, Tpt et Harmonia Amanda, *Atelier de formation à Wikidata*

REMERCIEMENTS

- Un grand merci à l'OIF pour avoir donné des bourses à certains des auteurs de ce travail tout en ayant la conscience que les langues de France et particulièrement l'arabe maghrébin constituent des composantes du patrimoine français et par conséquent des caractéristiques importantes de la francophonie.
- Un grand merci à Wikimedia France pour son accueil chaleureux à tous les participants de la WikiConvention Francophone 2016.



IL Y A UN INTÉRÊT POUR LA MISE EN RELIEF DES LANGUES MINORITAIRES ET LOCALES DE FRANCE DANS LES PROJETS WIKIMÉDIA

Wikipédia en corse



WIKIPEDIA
Enciclopedia libata

Wikipédia en occitan



WIKIPÈDIA
L'enciclopèdia liura

Wikipédia en breton



WIKIPEDIA
An holloueziadur digor

Wikipédia en basque



WIKIPEDIA
Entziklopedia askea

L'ARABE MAGHRÉBIN EST CONSIDÉRÉ AUSSI COMME UNE LANGUE DE FRANCE

[PDF] La " darja", langue de culture en France

D Caubet - Hommes et migrations, 2004 - academia.edu

... 193-204. • [Banalisation salutaire d'une «langue de France»: l'arabe maghrébin-darja sur la scène culturelle française

d'études péda
rapport aux m

Cité 16 fois

D Caubet - Champs linguistiques, 2006 - cairn.info

Si l'on observe
vingtaine d'
linguistique

Cité 4 fois

Locuteurs de l'arabe maghrébin-langue de France: une analyse sociolinguistique des représentations, des pratiques langagières et du processus de transmission

A Barontini - 2013 - tel.archives-ouvertes.fr

Qui sont les locuteurs de l'arabe maghrébin (dans ses diverses variétés) en France et quelles sont leurs pratiques langagières? Étudier ces pratiques suppose que l'on s'intéresse aux phénomènes migratoires liés à la (post) colonisation du Nord de l'Afrique. ...

Cité 5 fois [Autres articles](#) [Les 5 versions](#) [Citer](#) [Enregistrer](#)

NÉANMOINS,

Tunisian Arabic Wikipedia

This is an open test wiki of the Wikimedia Incubator.

Overview

There are:

- 95 articles, categories, templates, and talk pages;
- 48 redirects;
- 33 editors (including any with at least one edit);
- 478 revisions (including 84 minor edits).

Algerian Arabic Wikipedia

This is an open test wiki of the Wikimedia Incubator.

Overview

There are:

- 230 articles, categories, templates, and talk pages;
- 21 redirects;
- 62 editors (including any with at least one edit);
- 1549 revisions (including 97 minor edits).

Moroccan Arabic Wikipedia

This is an open test wiki of the Wikimedia Incubator.

Overview

There are:

- 184 articles, categories, templates, and talk pages;
- 58 redirects;
- 62 editors (including any with at least one edit);
- 971 revisions (including 193 minor edits).

DES PROBLÈMES...

Des dilemmes lors
du choix du script
à utiliser malgré
les
standardisations
déjà faite par
Caubet

- Respect du lien avec l'arabe moderne standard au niveau morphologique (Conventional Orthography for Maghrebi Arabic)

- Convertibilité vers la graphie arabe (Graphie latine de Turki et al.)

- Panarabisme (Interprétation politique de l'utilisation de la graphie latine)

Graphie arabe

- Simplicité de l'utilisation pratique (Graphie arabe de Caubet)

Graphie latine

- Simplicité de l'utilisation pratique (Graphie latine de Caubet, Arabizi de Caubet)

DES PROBLÈMES...

Un manque considérable de la reconnaissance du rôle et de la valeur de l'arabe maghrébin au Maghreb et en France malgré la dissémination significative de ce dialecte dans la société française et bien que « ces darijas » sont les variétés linguistiques les plus répandues du Maghreb

[PDF] [La transmission de l'arabe maghrébin en France: état des lieux](#)

A Barontini, D Caubet - Migration et plurilinguisme en France, 2008 - academia.edu

... Some case studies have shown that **Maghrebin** Arabic is still transmitted in some ... CAUBET Dominique 2007b, «L'**arabe maghrébin**-darja, une langue ressource en France», in Lambert ... CAUBET Dominique 2003, «La reconnaissance de l'**arabe** "dialectal" en France: un **parcours** ...

Cité 11 fois [Valorisation des langues vivantes en France: le cas de l'arabe maghrébin](#)

A Barontini - Le Français aujourd'hui, 2007 - cairn.info

... de l'ancienne épreuve ne savaient pas lire les caractères **arabes** et choisissaient ... D. (2003), « La reconnaissance de l'**arabe** "dialectal" en France : un **parcours sinueux** », dans J ... Le cas de l'**arabe maghrébin** », Communication à la journée Observer les pratiques linguistiques ...

Cité 5 fois [CITATION] [La reconnaissance de l'arabe "dialectal" en France: un **parcours sinueux**](#)

C Dominique - Lentin J. et L. Antoine (Éds.), Mélanges David Cohen, ..., 2003

Cité 9 fois [Autres articles](#) [Citer](#) [Enregistrer](#)

DES PROBLÈMES...

Vu leur utilisation excessive des emprunts du français et d'autres langues, les médecins et les professionnels du Maghreb trouvent de la difficulté pour expliquer leurs tâches pour le grand public bien qu'un jargon scientifique pour faire cela existe dans l'arabe maghrébin.

The language situation in Tunisia

M Daoud - Current Issues in Language Planning, 2001 - Taylor & Francis

This monograph describes the current language situation in Tunisia while maintaining a historical perspective that is helpful in understanding how language-related changes have come about, and a prospective view which may illuminate future developments. Even ...

Cité 40 fois

[Les attitudes professionnelles humaines des médecins de la région sanitaire de Sousse \(Tunisie\)](#)

F Daouas, AB Abdelaziz, TN Ajmi, A Mtiraoui - Santé publique, 2002 - cairn.info

Summary The new chronic and multi-factorial morbidity, the limited efficiency of classical medicine, as well as the preponderant position currently occupied by an individual in social life have imposed a patient-oriented approach as an ideal model of medical practice. The ...

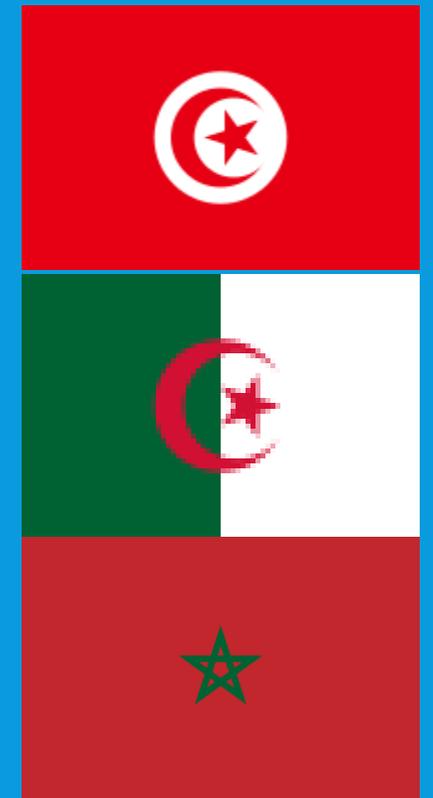
Cité 1 fois [Autres articles](#) [Les 3 versions](#) [Citer](#) [Enregistrer](#)

CE PROJET EST UTILE POUR...

L'amélioration
de la
couverture d'un
aspect
important du
patrimoine
français riche
et **varié** dans les
projets
Wikimédia



L'amélioration
de la couverture
des
compartiments,
des **civilisations**
et des
traditions du
Maghreb dans
les projets
Wikimédia



CE PROJET EST UTILE POUR...

L'étude automatisée
des particularités
morphologiques et
sémantiques de
chacun des
dialectes
maghrébins

The Right Honourable
John Curtin



14th Prime Minister of Australia
Elections: 1937, 1940, 1943

Personal details

Born John Joseph Curtin
8 January 1885
Creswick, Colony of Victoria
British Empire

Died 5 July 1945 (aged 60)
Canberra, Australia

per: date_of_birth per: city_of_birth

John Curtin (1885-1945),
prime minister and journalist,
was born on 8 January 1885
at Creswick, Victoria,
eldest of four children of
Irish-born parents John
Curtin and his wife
Catherine (Kate) Agnes,
née Bourke. ...

(Hirschberg &
Manning, 2015)

Learn how to extract
slots from free text:

was born on <DATE>

per: date_of_birth

born ... at <LOCATION>

per: city_of_birth

CE PROJET EST UTILE POUR...

L'amélioration de la culture générale des maghrébins de France et des français d'origine maghrébin qui ne maîtrisent très bien que l'arabe maghrébin

L'arrêt de la mauvaise interprétation des informations fournies par Wikipédia et projets associés causée par l'existence de plusieurs faux amis et de différences morphologiques significatives entre les dialectes maghrébins et l'arabe moderne standard.

Par exemple,

بندق signifie noisette dans l'arabe littéral et pignon de pin dans l'arabe tunisien.

Noisette est dite بوفريوة en arabe tunisien.

CE PROJET EST UTILE POUR...

La mise en commun des termes techniques (plantes, oiseaux, matériaux, maladies...) qui sont utilisés en arabe maghrébin et qui existent dans un nombre considérable d'ouvrages et par conséquent l'amélioration de l'intelligibilité du discours scientifique pour le grand public du Maghreb

[CITATION] **plantes** dans la médecine traditionnelle tunisienne

MK Boukef - 1986 - agris.fao.org

... **plantes** dans la médecine traditionnelle tunisienne. 1986. Boukef, MK. Translate with Translator. This translation tool is powered by Google. AGRIS and FAO are not responsible for the accuracy of translations. Related information in other data sources. ...

Cité 129 fois **Female issues in Arabic dialects: Words and expressions related to the female body and reproduction**

V Ritt-Benmimoun, S Procházka - 2011 - reuredc.uca.es

With contributions by: Rima Aldoukhi (Vienna), Sonia Aldulayme (Vienna), Alexandrine Barontini (Paris), Amina Bensalah (Paris), François Dumas (Aix-en-Provence), Marie-Aimée Germanos (Paris), Narine Grigoryan (Vienna), Amira Jaafar (Vienna), Leila Kaplan (Vienna), Hanan ...

Cité 2 fois [Autres articles](#) [Les 2 versions](#) [Citer](#) [Enregistrer](#)

C'EST POURQUOI, ON APPELLE...

الإسكوا национален שפה 文化 شرقية
i n a l c o

Institut national
des langues
et civilisations orientales

Alphatis

Association Alphatis

Cours d'arabe maghrébin à Paris



Délégation générale à
la langue française et
aux langues de France

épreuves du bac

22 "dernière minute" : les chiffres d'inscrits pour 1999 sont les suivants : arabe algérien 3191; arabe marocain 4817; arabe tunisien 1878; total 9886, soit une augmentation de 942 candidats (10,55%).

... POUR SOUTENIR NOS PROJETS DE LOCALISATION DE WIKIPÉDIA ET DE WIKTIONNAIRE...

- **Wiktionnaires (en arabe tunisien, en arabe algérien et en arabe marocain):** Chaque page contient les définitions du mot en dialecte arabe et en français, l'IPA du mot (à la manière des autres wiktionnaires en langues de France), le fichier sonore du mot (à la manière de Lingua Libre) et sa nature grammaticale.
- **Wikipédia en une variété standardisée de l'arabe maghrébin:** Similaire à la version écrite en anglais simple, elle peut être obtenue en traduisant les résumés des articles de la Wikipédia française (en utilisant l'outil de traduction) puisque chaque page est constitué d'un simple paragraphe munie de références (comme dans certaines versions de Wikipédia particulièrement Waray-Waray). Elle est écrite en graphie latine pour être intelligible aux maghrébins à l'étranger.
- **Wikipédias (en arabe tunisien, en arabe algérien et en arabe marocain):** À la manière de la Wikipédia écrite de l'arabe maghrébin standard mais utilisant la graphie arabe. Cette graphie arabe peut être purement phonologique ou en corrélation avec les règles orthographiques de l'arabe moderne standard.

..., POUR DISSÉMINER DES INFORMATIONS UTILES À PROPOS DE LA PARTICULARITÉ ET DE LA VALEUR LINGUISTIQUE DES DIALECTES MAGHRÉBINS DANS LA WIKISOURCE FRANÇAISE...

- Puisqu'il y a un manque significatif de reconnaissance de la valeur et de la particularité linguistique des dialectes maghrébins, on va publier les ressources linguistiques parlant de cela écrites avant 1923 et dont les auteurs sont décédés il y a plus que 70 ans sur la Wikisource française.
- Cela sera fait
 - En cherchant et ajoutant sur les versions numérisées de ces livres qui existent en ligne sous la forme de pdf
 - En créant la version wiki de ces travaux en utilisant l'outil OCR du site ainsi que les contributions des utilisateurs.



WIKISOURCE

...ET DANS LA WIKIPÉDIA ANGLAISE ET FRANÇAISE, ...

- Pour ces mêmes raisons, on a établi et on établit encore des articles dans la Wikipédia anglaise et française dans lesquels on explique la phonologie, la morphologie, l'histoire, l'utilisation et la sémantique des dialectes maghrébins tout en choisissant un style simple et intelligible aux utilisateurs.



ET POUR SOUTENIR NOS EFFORTS POUR L'IMPLEMENTATION DES DIALECTES MAGHRÉBINS DANS WIKIDATA...

- Puisque les données incluses dans toutes les wikipédias sont insérées dans WikiData, on a décidé lors de l'initiative de Mr. Helmi Hamdi de Wikimedia Tunisie de traduire WikiData en chacune des dialectes arabes du Maghreb en traduisant:
 - Les étiquettes des 24 millions d'éléments (oiseaux, plantes, termes scientifiques, objets...) qui existent dans WikiData en arabe tunisien, arabe algérien et arabe marocain.
 - Les messages systèmes de WikiData en arabe tunisien, arabe algérien et en arabe marocain en utilisant l'outil TranslateWiki
- Ainsi, on obtiendra sans aucune difficulté les noms de tous les termes techniques et les objets matériels du monde entier en dialectes maghrébins. Cet aperçu peut être utile par la suite pour
 - la création automatisée des Wikipédias en dialectes arabes maghrébins et des dictionnaires techniques bilingues par des bots
 - Améliorer l'efficacité des logiciels de traduction automatique de et vers l'arabe maghrébin
 - pour la dissémination du savoir scientifique avancé dans les dialectes arabes maghrébins dans les années à venir.
 - Promouvoir l'utilisation d'un lexique dialectal par les professionnels pour expliquer leurs travaux.

Item [Discussion](#) [Read](#) [View history](#)

Tunisian (Q56240)

Maghrebi dialect of the Arabic language spoken in Tunisia [edit](#)
Tunisian Arabic | ISO 639-3:aeb | Tunisian Darija | aeb

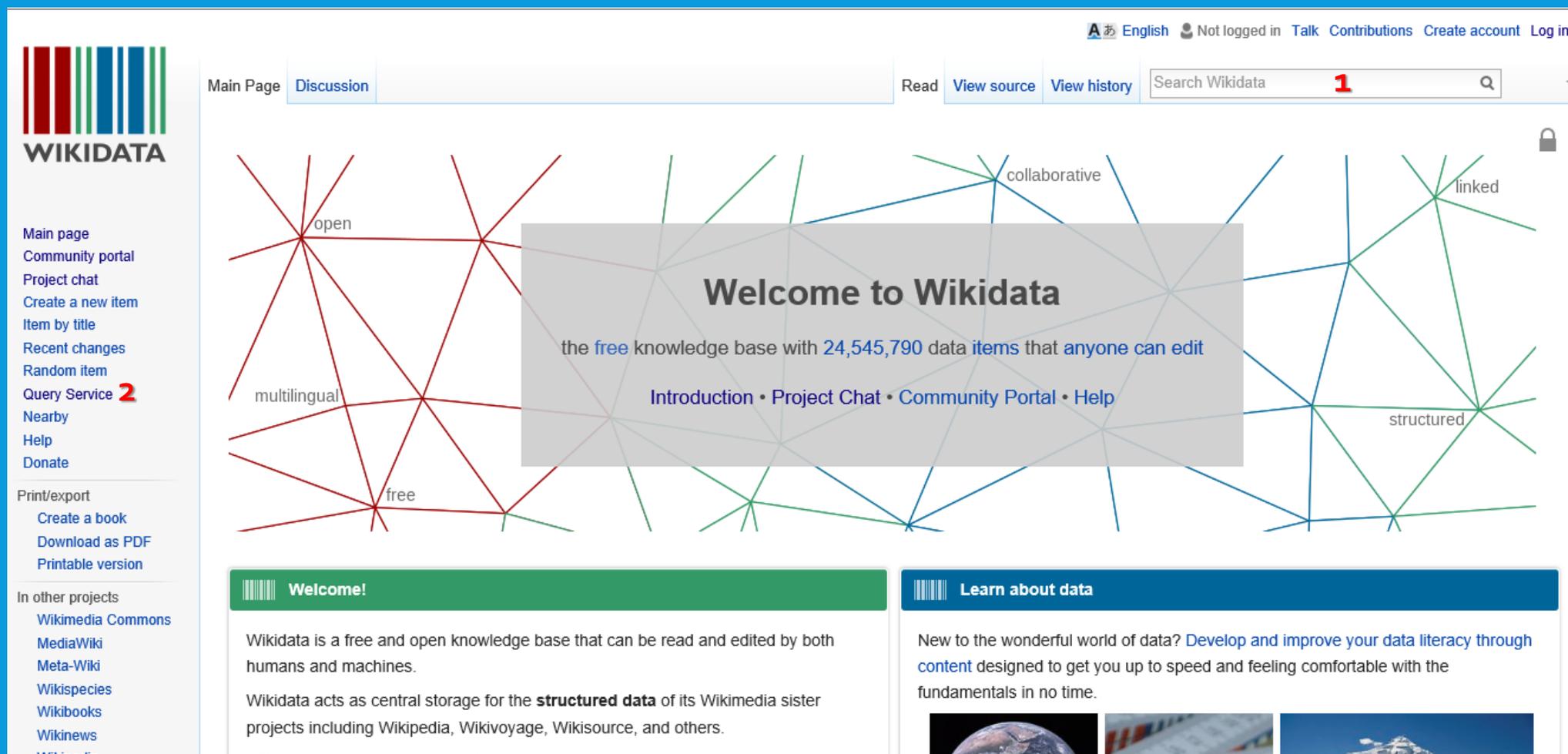
[In more languages](#) [Configure](#)

Language	Label	Description	Also known as
English	Tunisian	Maghrebi dialect of the Arabic language spoken in Tunisia	Tunisian Arabic ISO 639-3:aeb Tunisian Darija aeb
French	Tunisien	Dialecte maghrébin de la langue arabe, parlé en Tunisie	ISO 639-3:aeb Arabe tunisien
تونسي	التونسي	اللهجة العربي متاع تونس	دارجة تونسية
Algerian Arabic	No label defined	No description defined	

[All entered languages](#)

Comment chercher WikiData

1. Taper directement le nom du terme en anglais
2. Utiliser une requête SPARQL pour la recherche d'un terme inconnu en anglais en indiquant ses propriétés



The screenshot shows the Wikidata main page. At the top right, there are links for language (English), user status (Not logged in), and actions (Talk, Contributions, Create account, Log in). Below this is a navigation bar with 'Main Page' and 'Discussion' tabs, and a search bar containing 'Search Wikidata' and a red '1' icon. The main content area features a network diagram with nodes and connecting lines. A central grey box contains the text: 'Welcome to Wikidata', 'the free knowledge base with 24,545,790 data items that anyone can edit', and links for 'Introduction', 'Project Chat', 'Community Portal', and 'Help'. The network diagram includes labels like 'open', 'multilingual', 'free', 'collaborative', 'linked', and 'structured'. On the left side, there is a sidebar with various navigation options: 'Main page', 'Community portal', 'Project chat', 'Create a new item', 'Item by title', 'Recent changes', 'Random item', 'Query Service 2', 'Nearby', 'Help', 'Donate', 'Print/export' (with sub-options: 'Create a book', 'Download as PDF', 'Printable version'), and 'In other projects' (with sub-options: 'Wikimedia Commons', 'MediaWiki', 'Meta-Wiki', 'Wikispecies', 'Wikibooks', 'Wikinews', 'Wikipedia'). At the bottom, there are two boxes: 'Welcome!' and 'Learn about data'.

PLUS D'INFORMATIONS SUR L'ARABE MAGHRÉBIN ET SUR NOS PROJETS WIKIMÉDIA...

- Caubet, D. (1999). Arabe maghrébin: passage à l'écrit et institutions. *Faits de langues*, 7(13), 235-244.
- Caubet, D. (2000). *Propositions concernant la notation usuelle de l'arabe maghrébin: graphie arabe et graphie latine*. Paris: INALCO.
- Caubet, D. (2002). Arabe maghrébin, langue de France: entre deux graphies. *Les langues de France et leur codification écrite: écrits divers, écrits ouverts*, 307-330.
- Caubet, D. (2003). La reconnaissance de l'arabe "dialectal" en France: un parcours sinueux. In J. Lentin, & L. Antoine, *Mélanges David Cohen, Études sur le langage, les langues, les dialectes, les littératures, offertes par ses élèves, ses collègues, ses amis, présentées à l'occasion de son anniversaire* (pp. 135-148). Maisonneuve et Larose.
- Caubet, D. (2006). Banalisation salutaire d'une «langue de France»: l'arabe maghrébin-darja sur la scène culturelle française. *Champs linguistiques*, 161-168.
- Cerquilini, B. (2003). *Les langues de France*. Paris: PUF.
- Wikimedia Foundation. (2016). *Wikipedia, the Free Encyclopedia*.
- Wikimedia Foundation. (2016). *Meta Wiki*.

PLUS D'INFORMATIONS SUR L'ARABE MAGHRÉBIN ET SUR NOS PROJETS WIKIMÉDIA...

- Adel, E. (2016). *Standard Written Maghrebi*. Wikimedia Algeria.
- Zesch, T., Müller, C., & Gurevych, I. (2008). Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary. *LREC*, 8, pp. 1646-1652.
- Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, 349(6245), 261-266.
- Maamouri, M. (1998). *Language Education and Human Development: Arabic Diglossia and Its Impact on the Quality of Education in the Arab Region*.
- Maamouri, M. (1999). Literacy in the Arab region. *Literacy: An international handbook*, 400-404.
- Turki, H., Adel, E. (2016). L'arabe maghrébin standard. *WikiConvention Francophone*. Paris, France: Wikimédia France.
- Turki, H., Daouda, T., Adel, E., Regragui, N., Zribi, R., & Kerbouche, R. (2015). *Maghrebi Arabic Transcription Method*. *Wikipedia and Wiktionary Convention*. GRIN Verlag.
- Vrandečić, D., & Krötzsch, M. (2014). Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10), 78-85.
- Vrandečić, D. (2012, April). Wikidata: A new platform for collaborative data collection. In *Proceedings of the 21st International Conference on World Wide Web* (pp. 1063-1064). ACM.

AFFILIATIONS

- ^a Faculty of Medicine of Sfax, University of Sfax, Sfax, Tunisia
- ^b Société Nationale des Transports Ferroviaires, Algiers, Algeria
- ^c IGPro, Mostaganem, Algeria
- ^d Sbikha 1979 High School, Sbikha, Kairouan, Tunisia
- ^e University of Stockholm, Stockholm, Sweden
- ^f École de droit, Université Paris 2 Panthéon-Assas, Paris, France
- ^g Université Paris-Sorbonne, Paris, France
- ^h O-Maroc, Ouarzazate, Morocco
- ⁱ IESEG School of Management, Lille, France
- ^j Institut de Recherche en Immunologie et en Cancérologie, Université de Montréal, Montréal, Québec, Canada
- ^j Higher School of Economics, Peter the Great Saint Petersburg Polytechnic University, Saint Petersburg, Russian Federation
- ^k Université Hassiba Benbouali de Chlef, Hay Salem, Chlef, Algeria
- ^l Information Technology Office, United Nations, New York, United States of America
- * Wikimedia TN User Group, Sayada, Monastir, Tunisia
- ^ Wikimedia Algeria, Mostaganem, Algeria
- + Wikimedia Morocco, Salé, Morocco

ANNEXES

Pour ceux qui ne savent pas le déroulement de la création des wikis en arabe maghrébin

ANNEXE A: ÉTAPES DE LA CRÉATION D'UN WIKI EN DIALECTE MAGHRÉBIN

1 an

- Proposition du projet sur Meta-Wiki
- Discussion du projet avec la communauté WMF

2 ans

- Définition de la structure du wiki sur Incubator (Page d'accueil, Rubriques...)
- Développement des modèles à utiliser sur Incubator (Infobox...)

3 ans

- Résolution des problèmes techniques hors ligne (Manque d'utilisateurs intéressés, Manque de soutien, Problèmes de graphie, problèmes d'algorithme et de programmation, Problèmes de lexique...)
- Traduction des messages systèmes sur Translatewiki



WIKIMEDIA
META-WIKI



WIKIMEDIA
INCUBATOR



2 ans

- Ajustement de la structure finale du wiki dans Incubator pour inclure les solutions retrouvées pour les problèmes techniques
- Création d'articles pour la prolifération du wiki sur Incubator



2 ans

- Le projet est évalué par un spécialiste de la langue utilisée désigné par la comité linguistique de Wikimedia Foundation
- La comité linguistique de Wikimedia Foundation évalue le projet de point de vue rigueur et utilité et l'approuve



ANNEXE B: UTILISATEURS RESPONSABLES DES PROJETS WIKIMÉDIA EN ARABE MAGHRÉBIN

Projet Wikimédia	Utilisateurs responsables
Wikipédia en arabe tunisien	Csisc, Tounsimentounes, GeekEmad, E3
Wikipédia en arabe algérien	Oldstoneage, Reda Kerbouche, Bachounda, Gêt Fenyan
Wikipédia en arabe marocain	Anass Sedrati, Enzoreg, Dalinanir
Wikipédia en arabe maghrébin standard	GeekEmad
Wiktionnaire en arabe tunisien	Csisc, GeekEmad
Wiktionnaire en arabe algérien	Bachounda, GeekEmad

ANNEXE C: LES CODES ISO 639-3 DES PRINCIPAUX DIALECTES ARABES MAGHRÉBINS PRIS EN CONSIDÉRATION

Dialecte	Code ISO 639-3
Arabe tunisien en Alphabet du chat	AEB-LATN
Arabe tunisien en CODA tunisien	AEB-ARAB
Arabe algérien en graphie arabe	ARQ
Arabe marocain en graphie arabe	ARY