



Outillage de l'accès aux textes par la lecture active étymologique multilingue pour apprenants berbérophones et arabophones

Slimane Abdellaoui, Valérie Bellynck, Mathieu Mangeot, Christian Boitet

► To cite this version:

Slimane Abdellaoui, Valérie Bellynck, Mathieu Mangeot, Christian Boitet. Outillage de l'accès aux textes par la lecture active étymologique multilingue pour apprenants berbérophones et arabophones. Traitement Automatique des Langues Africaines TALAf 2018, Sep 2018, Grenoble, France. hal-02054881

HAL Id: hal-02054881

<https://hal.science/hal-02054881>

Submitted on 2 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Outillage de l'accès aux textes par la lecture active étymologique multilingue pour apprenants berbérophones et arabophones

Slimane Abdellaoui¹ Valérie Bellynck² Mathieu Mangeot³ Christian Boitet³

(1) Université Paris 13 et LIG-GETALP

(2), (3) Université Grenoble Alpes et LIG-GETALP

Slimane_abdellaoui@yahoo.com, {valerie.bellynck, mathieu.mangeot, christian.boitet}@imag.fr

RÉSUMÉ

Dans le cadre du projet Etymolo du LIG-GETALP, nous nous intéressons à la lecture active pour apprenants berbérophones, scolarisés en arabe, puis en français, puis étudiant le berbère de façon formelle. Dans ce type de situations, une aide de type *lecture active* peut être très utile, et elle devrait être *multilingue*. Comme il y a beaucoup de "cognats" (quasi-synonymes ou faux amis) entre ces langues, à l'écrit ou à l'oral, à cause de liens étymologiques ou d'emprunts, nous proposons d'aider les apprenants dans leur apprentissage lexical en leur montrant les cognats des mots du texte qu'ils lisent, arrivant ainsi à un outil de *lecture active étymologique* (LAEtymo). Nous avons recensé les aides apportables dans une dizaine de situations précises, et spécifié une extension de l'interface du CESSSELIN/LECTURE (dédié au japonais-français) pour montrer les équivalents lexicaux dans 1 à 4 langues cibles, et pour intégrer l'aspect "étymologique" de façon qu'il constitue une aide significative. À partir de l'outil CESSSELIN/LECTURE, nous avons produit une simulation de cette interface, avec en source le berbère et en cible le français, le berbère, l'arabe et l'arabe dialectal algérien. Comme il s'agit de couples de langues peu dotés, notre premier problème est de collecter ou constituer les ressources nécessaires, à savoir un grand ensemble de cognats entre ces langues, une base lexicale utilisable pour chaque couple de langues, et un lemmatiseur pour chaque langue. Dans le cas du berbère, nous sommes en train de collecter des dictionnaires et de les mettre sur la plate-forme JIBIKI, ainsi que des corpus monolingues en tamazight (écrits en tfinagh ou tamamritrit —latin), le tout en format XML et en encodage UTF-8.

ABSTRACT

Helping access to texts through etymological multilingual active reading for learners of Berber and Arabic

In the context of the Etymolo project of LIG-GETALP, we are interested in active reading for Berber speakers, schooled in Arabic, then in French, then studying Berber formally. In this type of situations, a help like *active reading* can be very useful, and should be *multilingual*. Since there are many "cognates" (complete synonyms or words spelt in the same manner but not having the same meaning) between these languages, as spoken or written words, because of etymological relationships or borrowed words, we propose to help learners in their lexical learning by showing them the "cognates" of the words in the text they read, coming then to the concept of *etymological active reading* (LAE). We identified the possible helps in ten real situations and specified an extension of the CESSSELIN/LECTURE interface (dedicated to Japanese-French) that will show the lexical equivalents in 1 to 4 target languages, and include the "etymological" aspect, so that it constitutes a significant help. On the basis of CESSSELIN/LECTURE, we have produced a simulation of this extended interface, with Berber as source, and French, Arabic and Algerian Arabic dialect as targets. Since we handle under-resourced language pairs, our first problem is to collect or constitute necessary resources, that is a large set of cognates between these languages, a lexical basis usable for each language pair, and a lemmatizer for each language. In the case of Berber, we are collecting dictionaries and putting them in the JIBIKI platform, as well as monolingual corpora in Berber (written in tfinagh or tamamritrit — latin), all of these in XML format and UTF-8 encoding.

AGZUL

Asfecku n tezmareft yer yeḍrisen s tyuri turmiṭ tagmuẓart tagtutlayant I nelmaden imaziyen d w aeraben

Deg wahil n usenfar Etymolo n LIG-GETALP, tamuṭli nney twehha yer leqraya turmiṭ tigertutlayant I yinelmaden imaziyen i ylemden taerabt, tarumit, sakn ad nwali tamaziyt s lwehla taylant. Di teswiṭ icuban tagi, lemeawna am tyuri turmiṭ tezmer ad tili d ayen inefeen s waṭas, ilaq ad tili tagtutlayant. Imi llan watas n wawalen yettemcabin ger tutlayin agi, akken ittwanṭaqen ney i twarun yef demma n yissayen igmuẓar, nesummer-d aneiwen inelmaden di leqraya nsen tamawalant mi ara sen nesken awalen itemcabin mi ara ten-nlemden, s wayagi ad naweḍ yer leqraya turmiṭ tagtutlayant. Nwala lemeawnat deg watas n teswiein tissediyn d usnarni n ugrudem n CESSSELIN/LECTURE (i tɣapunit d tfransist) iwakken ad nesken igdazalen imawalen ger n yiwet d wukuz n tutlayin dayen ad as-d-nernu "tagmuẓart" iwakknen ad tɛiwen akken iwata. Gef lsas n Cesselin/lecture nebna-d asimey n ugrudem agi, tamaziyt d taybalut, taerabt tantala d tfransist d tinicanin. Imi d ayen icudden yer tutlayin ur yesein ara tawilat meqren, aybel nney amenzu mi ara d-nettnadi ney ad nebnu tiybula ilaqen, ger-asen atas n wawalen itemcabin ger tutlayin, azadur amawalan ara yettwaqedcen i yal tutlayt. Aqlay nettheyyi isegzawalen i tutlayt n tmaziyt iwakken ad tt-ner di teyeryert Jibiki, arnu-yas ugruwen aynutlayen s tmaziyt (s tfinay ney s tmamrit) s umasal XML d ustengel UTF-8.

MOTS-CLÉS : Lecture active, lecture active étymologique, apprentissage lexical, apprentissage des langues, berbérophonie.

KEYWORDS: Active reading, etymological active reading, lexical learning, language learning, berberophony.

Introduction

Le berbère est une macrolangue¹ peu dotée parlée en Afrique du Nord par environ 45 millions de personnes. Elle contient plusieurs variantes, chacune ayant son propre lexique qui la distingue des autres, ainsi que des différences de flexion. Les locuteurs de cette langue l'étudient dans des écoles, généralement en parallèle avec l'arabe et le français. Ils passent d'une langue à une autre, et les mots qui se ressemblent, à l'écrit ou à l'oral, leur facilitent l'apprentissage du lexique de ces langues. Nous les appellerons cognats², même s'il s'agit de "faux amis" (comme "dewan", bureau, et "divan", canapé).

La lecture active est une méthode utilisée pour faciliter la compréhension de textes pendant la lecture en ligne. La Figure 1 donne un exemple tiré de <http://laosoftware.com/>, où le texte est présenté verticalement, un mot par ligne, avec en regard la prononciation et les équivalents possibles en français.

Aide à la lecture active

ນິທານເລື່ອງນີ້ກໍ່ຄວາມສະເຫຼີອນໃຈແກ່ຜູ້ຄົນແຕ່ບູຮານນະການມາແລ້ວເລື່ອງເລີ
ມຕົ້ນຂຶ້ນເມື່ອເມື່ອຍື່ງສອງຄົນພາກັນໄປອາບນ້ຳຢູ່ແຄມທ່ານ້ຳຫ້ວຍໃສ່ໄຫວເຍ້ນຈົນຫວຽວ
ເຫັນໝູ່ປາບາງຄັ້ງກໍ່ມີໃບໄມ້ແຫ້ງຫລິ້ນໄຫວມາຕາມໜ້ານ້ຳ

Nouvelle traduction

Traduction mot à mot du texte

(les traductions se basent sur un dictionnaire aimablement fourni par Paul Jadin)

ນິທານ (nithan) : [fable](#) / [conte](#) / [récit](#)
ເລື່ອງນີ້ (lūangni) : *forme proche ou ancienne (ເລື່ອງນີ້)* : [sujet](#) (à ce -)
ກໍ່ (ko) : [construire \(en dur\)](#) / [bâtir](#) / [commencer](#) / [débuter](#)
ຄວາມ (khwaam) : [mot](#) / [idée](#) / [parole](#) / [argument](#) / [proposition](#) / [terme servant à substantiver un verbe](#)
ສະເຫຼີອນ (sathuan) : [trépider](#) / [secoué \(être\)](#) / [palpiter](#) / [tremblant](#) / [palpitant](#)
ໃຈ (chay) : [cœur](#) / [sein](#) / [pensée](#) / (mot concernant la pensée...)
ແກ່ (kè) : [à l'égard](#) / [à \(envers\)](#) / [vieux](#) / [ancien](#) / [âgé](#) / [mûr](#) / [expérimenté](#) / [foncé](#) / [sombre](#) / [doyen \(notable\)](#) / [conseiller](#) / [ancien \(nom\)](#)
ຜູ້ (phou) : [personne](#) / [celui qui](#) / [mâle](#) / [masculin](#)
ຄົນ (khon) : [personne \(terme générique:quelqu'un\)](#)
ແຕ່ບູຮານນະການ (tèbouhannakan) : [depuis l'antiquité](#)

Figure 1 : Lecture active "verticale" pour le lao-français

La lecture active étymologique multilingue n'est pas encore utilisée dans ce domaine. Dans le cadre du projet *Étymolo* du LIG-GETALP, on souhaite étendre la lecture active existante à la lecture active étymologique (utilisation des cognats), et à la rendre multilingue. Notre travail sur le berbère est le premier essai de progresser dans cette voie.

Le plan de cet article est le suivant. La section 1 contient la présentation du projet et du contexte : le principe de la lecture active étymologique, le contexte local pour le berbère, la coexistence entre le berbère, l'arabe et le français en Afrique du Nord, des exemples, et différentes situations d'apprentissage. La section 2 concerne les situations d'utilisation. La section 3 présente le maquetage de l'interface souhaitée, les méthodes et les étapes suivies, les outils utilisés, et la méthode de collecte de ressources pour réaliser le projet.

1 Présentation du projet et du contexte

1.1 Le projet *Étymolo* en général

Citons (à l'aide d'une fonte spéciale) la présentation du projet LIG-émergence *Etymolo* par V. Bellycnk et M. Mangeot (2017).

Le projet *Étymolo* vise à faciliter la lecture aux apprenants d'une langue en établissant des liens avec le lexique d'une ou plusieurs langues qu'ils connaissent aussi, en utilisant l'analogie³ (ou la simple similarité écrite ou orale) des mots de langues voisines. Pour cela,

¹ Voir <https://fr.wikipedia.org/wiki/Macro-langue>

² Voir https://fr.wikipedia.org/wiki/Mot_apparenté

³ L'analogie est une relation quaternaire : "A:B :: C:D" $\equiv$ "A est à B ce que C'est à D". Voir le HDR de Y. Lepage (2003). La ressemblance est une relation binaire symétrique. La similarité est une mesure telle que : $\text{sim}(A,B) = \text{sim}(B,A) \in [0, 1]$ et $\text{sim}(A, A) = 1$.

on désire créer un environnement informatique à partir de l'outil CESSSELIN/LECTURE, qui permet la lecture active du japonais en français, et est implémenté sur la plate-forme JIBIKI (Mangeot, 2014).

Voici un exemple de CESSSELIN/JIBIKI. On remarque que chaque mot est surmonté de sa prononciation, et qu'un article de dictionnaire JP-FR apparaît lors du survol d'un mot (s'il est dans le dictionnaire). Par contre, il n'y a qu'une seule langue cible (alors qu'on pourrait utiliser la base lexicale PAPILLON, voir <http://papillon-dictionaries.net/>) pour avoir en cible le français et l'anglais. Avec cette présentation, on ne peut pas, au contraire de la précédente, faire apparaître ensemble tous les articles de dictionnaire concernant les mots du texte source.



Figure 2 : Écran de lecture active japonais-français de Cesselin/lecture

1.2 Le projet Étymolo/berbère

Notre but est d'étendre cette plate-forme à d'autres langues comme l'arabe et le berbère en utilisant l'étymologie comme un moyen qui pourra augmenter les chances de mieux comprendre des textes des langues d'apprentissage.

Les traitements nécessaires à la réalisation de cette plate-forme sont principalement l'utilisation d'outils de traitement automatique des langues, dont des analyseurs morphologiques, ainsi que la consultation d'une base lexicale. Nous voulons étendre ce concept pour aider à retenir la forme des mots et faciliter l'intercompréhension dans des langues voisines. Il s'agit alors d'afficher des constituants des mots, soit pour leur origine étymologique, soit pour leur ressemblance morphologique ou phonologique.

La lecture active étymologique est un moyen qui peut non seulement faciliter la compréhension d'un texte écrit dans une langue en cours d'apprentissage (grâce aux informations qui s'affichent sur les mots du texte pendant la lecture), mais aussi faciliter la mémorisation des mots nouveaux rencontrés et de leurs sens. Présenter à l'apprenant(e) des "cognats"⁴, i.e. des mots déjà connus dans une ou plusieurs langues et qui ressemblent aux mots (encore) inconnus rencontrés, lui permettra de mieux comprendre le sens partagé (souvent celui de l'étymon), et surtout d'établir des liens mnémoniques. Prenons par exemple un arabophone qui connaît le français et commence à apprendre l'anglais. En lisant le texte, il tombe sur "autumn". L'idée est de lui présenter le cognat français "automne" en affichant simplement : *automne(fr)* = *autumn(eng)*.

1.3 Contexte concernant le berbère

Le berbère est un représentant des langues africaines moins dotées et actuellement en voie de développement. Les locuteurs berbères apprennent, en général, deux langues (arabe et français) durant leur parcours scolaire, et d'autres langues comme l'anglais, l'espagnol et l'allemand (le cas se présente en Algérie) arrivent par la suite.

D'après Wikipedia et d'autres⁵, l'amazigh (tamazight en berbère) est une "macrolangue" (comme le serbo-croate, ou

⁴<https://fr.wiktionary.org/wiki/cognat>

⁵<https://lavoixdelafemmeamazighe.weebly.com/le-peuple-amazighe.html>

l'hindoustani), dérivée du lybique ancien, qui regroupe un certain nombre de variantes, parlées par environ 45 millions de personnes en Afrique du Nord. Le touareg (Mali, Niger, etc.), le chleuh (Maroc), le kabyle (Algérie), etc. en font partie. On parle aussi des langues berbères. Mohand Mahrazi (Mahrazi, 2011) écrit que :

« ...la langue berbère constitue la langue autochtone de l'Afrique du Nord. Elle est parlée dans une dizaine de pays, et par des dizaines de millions de personnes sous forme de dialectes non standardisés et employée pour des besoins de communication essentiellement orale. »

Nous trouvons dans Wikipedia l'information suivante, légèrement adaptée.

Cette macrolangue a maintenant presque partout un statut officiel, et elle est étudiée comme une langue importante, en parallèle avec 2 ou 3 langues. Dans son aire de répartition, la langue maternelle est soit un dialecte arabe, soit un dialecte berbère.

L'école maternelle est presque toujours en arabe⁶, l'école primaire est le plus souvent en arabe, ou en français (école privée), le berbère est enseigné toujours à côté de ces deux langues, mais pas seule, puis l'école moyenne (collège) et secondaire est en français, et l'arabe ainsi que le berbère sont également enseignés de façon "formelle" (comme des matières d'enseignement, et au même titre ou presque que l'anglais et d'autres langues de grande diffusion comme l'espagnol, l'allemand, etc.).

Dans le contexte marocain, l'enseignement est en arabe dès la première année à l'école, et ces derniers temps l'amazigh (berbère) est aussi enseigné au même temps que l'arabe. Quant au français, il est enseigné à partir de la 3ème année du primaire. L'anglais est enseigné au secondaire.

Non seulement les jeunes actuellement scolarisés, mais aussi d'autres personnes sont en situation d'apprentissage du français ainsi que de leur dialecte berbère (par exemple en Algérie, où la variante dominante est le kabyle) et de l'arabe "classique" dans des écoles privées (écoles de langues), maisons de jeunes, associations, et même des classes dans le programme d'alphabétisation des adultes, etc.

1.4 Existence de cognats entre le berbère, l'arabe et le français

Une longue coexistence (Taleb Ibrahim, 2006) entre les trois langues (berbère, arabe et français) dans les pays d'Afrique du Nord a créé une relation linguistique étroite⁷. Les berbérophones ont beaucoup emprunté à l'arabe et les arabophones ont beaucoup emprunté au berbère ainsi qu'au français.

1.4.1 Cognats entre arabe dialectal et berbère

Il y a plusieurs mots utilisés en berbère qui sont identiques à ceux utilisés en arabe classique et en arabe dialectal, ou très proches d'eux. Inversement, il y a des mots dans l'arabe dialectal qui sont d'origine berbère.

L'exemple suivant montre la présence de cognats entre l'arabe classique, l'arabe dialectal algérien et le berbère ; le français est donné à titre informatif.

Arabe classique	يحتاجون للاستعمال اليومي لهذه الأدوية
Translittération	iahtāḡūna llisti'māl al-iaūmi ilhaḍiḥi al-adwya
Arabe algérien	يحتاجو لهاد الدواوات باش يستعملوها كل يوم
Translittération	ihtaḡu l haḍi dwawat bāš ista'mluha kul yūm
Berbère	Hwaḡen dwawi agi akken ad tsexdamen yal ass
Français	Ils ont besoin de ces médicaments pour l'usage quotidien

Tableau 1 : Exemple montrant plusieurs cognats à l'aide de couleurs

1.4.2 Cognats entre français et arabe

Dans le cas du couple français-arabe, il y a beaucoup de mots proches entre les deux langues, en particulier dans les textes scientifiques et dans les textes sur les nouvelles technologies, comme dans la phrase suivante.

Arabe classique	ثلاث عمليات أساسية تحدد النظام التقني المعقد "تلفزيون"...
Translittération	ṭalāt 'amliāt asāsia tuḥadid al-niẓām al-tiqnī al-mo'aqad "tilifiziūm"...
Français	Trois opérations fondamentales définissent le système technique complexe "télévision"...

⁶<http://www.ac-grenoble.fr/casnav/medias/pdf/scolarite-algerie.pdf>

⁷<https://journals.openedition.org/anneemaghreb/305?lang=ar>

Tableau 2 : Exemple de cognats entre l'arabe et le français

1.4.3 Cognats entre français et berbère

Le couple français-berbère a aussi beaucoup de cognats à cause des emprunts du berbère au français.

Voici deux exemples tirés de (Feraoun, 2004) et de (Allioui, 2007).

Berbère	lakul agi s ttiqan n djaj, s twabel timaynutin.
LA-Fr	école cette avec fenêtres en verre avec tables nouvelles
Français	Cette école avec des fenêtres en verre, et nouvelles tables
Berbère	Teryel t-ṭamhajt, ṭasufajt
LA-Fr	Ogresse inhumaine sauvage
Français	L'ogresse est inhumaine et sauvage

Tableau 3 : Deux phrases montrant la présence de cognats entre le berbère et le français

2 Situations d'utilisation

2.1 Situations typiques d'apprentissage

Il y a différentes situations d'apprentissage des langues dans le domaine d'extension des langues berbères, et une même personne peut en rencontrer plusieurs au cours de sa vie. Voici trois situations typiques d'un vécu courant en Algérie.

Ghilas est une personne née en Algérie du côté de Tizi Ouzou (Kabylie) ; pendant son enfance il a appris le kabyle. Ce dialecte du berbère est parlé par ses parents. C'est donc sa langue maternelle. À l'âge de 5 ans, il est entré à l'école primaire, où l'enseignement est en arabe, et il a compris assez facilement ce qu'il lisait dans le manuel, dès qu'il a appris l'alphabet arabe, parce qu'il a trouvé plusieurs mots qui sont proches (des emprunts) entre l'arabe et le berbère. À l'âge de 7 ans, il a commencé à apprendre le français. À 10 ans, il est entré au collège et il a commencé à apprendre l'anglais. À ce stade, il pouvait déjà comprendre plusieurs mots anglais pendant la lecture, grâce à leur ressemblance avec des mots français.

Amine est aussi né en Algérie du côté d'Oran ; il a appris pendant son enfance l'arabe dialectal. À l'âge de 5 ans, il est entré à l'école où l'enseignement est en arabe littéral, et il a compris facilement ce qu'il lisait en arabe car les mots sont très proches de ceux de l'arabe qu'il parle à la maison. À 7 ans, il a commencé à apprendre le français. Au début, il a rencontré des difficultés, mais quelques mois après il a réussi à comprendre mieux. Il a commencé à apprendre le tamazight à 8 ans. Il comprenait alors plusieurs mots pendant la lecture, car c'étaient des cognats, comme dans la phrase suivante (qui se trouve dans le livre de 4^{ème} année de primaire).

Berbère	Di zzman amezwaru, nnejmaɛen lewhuc
Arabe	في قديم الزمان / اجتمع الوحوش
Translittération	fī qadīm al-zmān iġtamaʿt al wuhūš
Français	Il était une fois, les animaux se sont réunis.

Tableau 4 : Phrase présentant des cognats en berbère et en arabe classique

Une troisième situation est vécue par de nombreux adultes. Ils apprennent des langues sur des sites Web, grâce à des environnements destinés à l'apprentissage des langues. Dans leur situation, il s'agit d'auto apprentissage pour lequel un outil de lecture active est très efficace. Pour chaque situation il nous convient d'offrir des aides appropriées à l'apprentissage de l'ensemble des langues étudiées.

2.2 Tableau de synthèse des situations et des aides possibles

Dans le tableau suivant, nous synthétisons différents scénarios possibles d'apprentissage des langues concernées, dans différents contextes, en donnant des exemples de cognats apparaissant entre les langues connues et les langues d'apprentissage.

Le tableau est organisé en 7 colonnes. la première colonne contient la langue maternelle de l'apprenant, la deuxième les langues déjà connues par l'apprenant, la troisième la langue que le locuteur commence à apprendre, la quatrième le nombre estimé d'apprenants concernés par la langue d'apprentissage, la cinquième le contexte d'apprentissage, la sixième des exemples tirés de textes (journaux, manuels scolaires, sites internet, etc.) des langues d'apprentissage et démonstration de cognats, enfin la septième concerne les ressources lexicales (Lemmatiseurs et dictionnaires) des langues de comparaison.)

Langue maternelle	Langues connues	Langue d'apprentissage	#	Contexte d'apprentissage	Exemples	Ressources utiles
berbère	arabe	français	2,5 M	Algérie, à partir de 3 ^{ème} année de primaire	Fr: La gazelle est terrifiée LA-Ar: الغزالة خائفة LA-Be: Tayzalt tugad LA-Ad: لغزالة خايفة Ar: الغزالة خائفة Ad: لغزالة خايفة Tr: Legzāla haifa Cognats: Fr: gazelle, Ar: الغزالة, Ad: لغزالة, Be: tayzalt	Lemmatiseur français, arabe Dictionnaires français - arabe arabe - français
berbère	arabe + français	anglais	1 M	Algérie à partir de la première année moyenne	En: I've got some tomato soup. LA-Fr: J'ai peu tomate soupe LA-Ar: خساء الطماطم بعض لّذي LA-Ad: لاسوب طماطيش شوية عندي LA-Be: seiγ cuya tomatic lasup Fr: J'ai un peu de soupe de tomate Ar: لّذي بعض خساء الطماطم Tr: Ladaia ba'ḍ ḥasā' al-ṭmāṭim Ad: عندي شوية لاسوب تاع طوماطش Tr: 'endī šuwy lāsūb(p) tā' tūmāṭiṣ Be: seiγ cuya n lasup n tomatic Cognat : En : soup, tomato, Fr: soupe tomate, Ar: طماطم, Ad: لاسوب, Be: lasup	Lemmatiseur français, anglais, arabe Dictionnaires anglais - français anglais- arabe
arabe	français	anglais	4 M	Algérie à partir de la première année moyenne	En: The table is in front of the sofa LA-fr: La table en face le sofa LA-Ar: الأريكة مقابل لطاولة LA-Ad: الفوتاي مقابل الطاولة Fr: La table est en face du sofa Ar: الطاولة مقابل الأريكة Tr: al-ṭāwila mqābil al-arika Ad: الطاولة مقابل الفوتاي Tr: ṭṭābla meqābel el-futai Cognats: En: table, sofa, Fr : table, sofa, Ar: الطاولة, Ad: الفوتاي, Ber: tabla, afutai	Lemmatiseur anglais, français Dictionnaires anglais - français anglais - arabe
berbère	arabe + français + anglais	espagnol	10K	Algérie 2 ^{ème} année secondaire (étudiants en langues étrangères)	Es: Es fue el algodón con alcohol LA-fr: Ce être le coton avec alcool LA-en: that to be cotton with alcohol LA-Ar: نالك كان قطن مع الكحول LA-Ad: هناك كان لفظن معا لالكول LA-Be: Winna d leqten ag-d lalkul Fr: C'était le coton avec de l'alcool En: It was cotton with alcohol Ar: كان ذلك قطن مع الكحول Tr: kāna ḍalika qūṭn ma'a al-kuḥūl Ad: كان لفظن معا لالكول Tr : k̄an leq̄ten m̄'a l̄alkūl Cognats: Es: algodón, alcohol, Fr : coton, alcool, En : coton, alcohol, Ar: قطن, كحول, Ad: لفظن, لالكول, Be: leqten, lalkul	Lemmatiseur espagnol, anglais, français, arabe Dictionnaires espagnol-anglais espagnol-français espagnol-arabe
arabe	arabe + français + anglais	allemand	10 K	Algérie à partir de la 2 ^{ème} année secondaire (étudiants en langues étrangères)	Al: Meine Familie besteht aus 4 Angehörigen LA-Fr: Ma famille être constitué de 4 ressortissant(s) LA-En: My family be_composed of 4 member(s) Fr: Ma famille est composée de quatre personnes En: My family is composed of 4 persons Cognats: Al:Familie, Fr: famille, En:family	Lemmatiseur allemand, anglais, français Dictionnaires Allemand-anglais Allemand-français

Langue maternelle	Langues connues	Langue d'apprentissage	#	Contexte d'apprentissage	Exemples	Ressources utiles
arabe	français + anglais	espagnol	20 K	Algérie 2 ^{ème} année au lycée (étudiants en langues étrangères)	Es: La contaminación en nuestra planeta es muy grave LA-fr: La pollution dans notre planète est très grave LA-En: The pollution on our planet is very dangerous LA-Ar: جد كوكب نا على التلوث خطير Fr: La pollution sur notre planète est très grave En: The pollution in our planet is very serious Ar: التلوث على كوكبنا خطير جدا Tr: al-talawuṭ ‘alā kawkabina ḥaṭir ġidā Cognats: Es: contaminación, planeta, ist, grave, Fr : contamination, est, grave, En: contamination, is, grave	Lemmatiseur espagnol, anglais, français Dictionnaires Espagnol-anglais Espagnol-français Espagnol-arabe
arabe	français	berbère	1M	Algérie à partir de la 4 ^{ème} année primaire	Be: Yelsa umyar abernus LA-Fr: vêtu le vieux burnous LA-Ar: برونوس شيخ لبس Fr: Le vieux est vêtu d'un burnous Ar: لبس الشيخ البرنوس Tr: labasa al-šaiḥ barnūs Ad لبس الشيخ برونوس Tr: lbes ših barnūs Cognats: Be: abernus, Fr: burnous, Ar: البرنوس, Ad: البرنوس	Lemmatiseur berbère, français dictionnaires berbère français
français	français	arabe	10 K	France dès le CP	Ar: كَانَتْ رَائِحَتُهُمْ مِثْلَ الْيَاسْمِينِ Tr: kānat rā'ihatahuna miṭla al iāsamin LA-Fr: était parfum-leur comme le jasmin Fr: Leur parfum était comme le jasmin Cognats: Ar: الياسمين, Ad: لياسمين, Fr: jasmin, Be: liasmin	Lemmatiseur Arabe, français Dictionnaire Arabe-français
anglais	anglais	berbère	200	Canada (Ottawa) Première année à l'école	Be: D acu id trisiti ? LA-En: Is what electricity En: What is electricity? Cognats: Be: Trisiti, En: Electricity.	Lemmatiseur berbère, français dictionnaires berbère-français

Tableau 5 : Différentes situations pour l'apprentissage des langues

2.3 Remarques sur le tableau

Dans la colonne "#", on trouve une estimation du nombre de locuteurs dans tous les domaines des langues berbères en Algérie, et pour les dernières lignes, en France (nous n'avons pas les chiffres pour la Belgique et la Suisse) et au Québec (nous n'avons pas les chiffres pour le Nouveau-Brunswick)⁸. En ce qui concerne les exemples, nous avons mis en premier la langue d'apprentissage, puis la lecture active des langues connues auparavant, puis la traduction normale.

En ce qui concerne les ressources, nous avons mentionné les dictionnaires bilingues et les lemmatiseurs des langues concernées dont on peut avoir besoin pour chaque langue selon la situation d'apprentissage.

3 Maquettage de l'interface souhaitée

3.1 Points de départ

3.1.1 CESSÉLIN/LECTURE pour japonais et français

Une personne dont le français est la langue maternelle, et qui apprend le japonais, rencontre, alors qu'elle voyage au Japon,

⁸ Les chiffres du tableau sont tirés de <http://countrymeters.info/fr/Algeria>, <http://fr.allafrica.com/stories/201704050949.htm>, <https://www.tsa-algerie.com/enseignement-de-tamazight-en-algerie-etat-des-lieux/>, <https://www.liberte-algerie.com/culture/nous-voulons-renforcer-l-enseignement-de-lespagnol-en-algerie-241033>

des difficultés à comprendre le contenu de textes en japonais. L'outil CESSSELIN/JIBIKI (Mangeot, 2014) pour le couple japonais-français apporte une aide de type lecture active.

Si on colle (ou saisit) un texte en japonais dans la zone prévue à cet effet, alors au survol de la souris, on voit la traduction du mot sur lequel on pointe le curseur. L'affichage est mot par mot (présentation horizontale). Au-dessus des parties de mots en kanji, on voit leur transcription phonétique en hiragana, en rōmaji (kunrei), ou en rōmaji (kunrei). Lors du survol d'un mot qui a une ou plusieurs entrées dans la base lexicale, la traduction des lemmes correspondants apparaît dans un article « reconstitué ».



Figure 3 : Capture d'écran de lecture active avec CESSSELIN/LECTURE/JIBIKI pour le japonais-français

3.1.2 SUFT-1/Jibiki (affichage permanent)

SUFT1/JIBIKI est un autre outil de lecture active multilingue pour les tweets (système d'aide à la compréhension pour des tweets étrangers). Son affichage est permanent, avec une présentation horizontale complexe ou une présentation verticale (Shah, 2017). Voir la Figure 4 ci-dessous.

3.1.3 Ressources pour le maquetage : phrases exemples, articles de dictionnaire et table de cognats (amazigh)

3.1.3.1 Phrases exemples

Nous avons choisi de nous concentrer sur le berbère et de construire la maquette à partir d'un petit nombre de phrases en berbère.

Nous avons choisi 3 phrases qui contiennent des cognats dans 10 textes en amazigh (écrits en tamazight) du journal "La Dépêche de Kabylie"⁹, et 2 phrases d'un texte parmi 5 écrits en tifinagh (sujet du bac blanc 2017, Office National des examens et concours¹⁰).

Après avoir compté les cognats dans 15 pages (5 en tifinagh et 10 en tamazight) d'environ 250 mots chacune, nous avons trouvé qu'il y a en moyenne 20 cognats (entre 15 et 25) par page.

Nous avons choisi des phrases courtes pour qu'elles tiennent en une ligne sur la maquette.

⁹ <http://www.depechedekabylie.com/>

¹⁰ <https://www.exoco-lmd.com/tamazight/sujets-bac-algerie-et-correction-tamazight/>



Figure 4 : Écran de lecture active japonais-anglais de tweets japonais avec SUFT-1

- P1. Be: Muqlet aselmed n tmaziyt, muqlet **tilibizyun** ney **radio**.
 Ar: أنظروا إلى تعليم الأمازيغية، شاهدوا التلفزيون أو اسمعوا الراديو
 Tr: unẓuru ila taʿlīm al-amāziġia, šāhidū al-tlfziūn aw ismaʿū al-rādiū
 Ad: شوفو تلفزيو ولا سمعوا الراديو
 Tr: šūfū leqraia ta3 tamāziġt šūfū **tilfiziū** wla smʿū **al-rādiū**
 Fr: Regardez l'enseignement du tamazight, regardez la **télévision** ou écoutez **la radio**.
- P2. Be: Aṭas n yimawlan i **yettlummun** arraw-nsen
 Ar: الكثير من الأولياء يلومون أو لادهم
 Tr: al-kaṭīr min al-awliāʾ **ialūmūna** awlādahum
 Ad: بزاف من لوالدين ايلوموا ولادهم
 Tr: bzāf men lwāldīn **ilūmū** wlādhūm
 Fr: Plusieurs parents blâment leurs enfants
- P3. Be: Imsunayen-a, ad sbegnen, di **tsebhīt**, ayen i sunnuyen
 Ar: هؤلاء الرسامون سيظهرون في الصبيحة رسوماتهم
 Tr: hauʿlāʾ al-rsāmūn saīuẓhrūna fī **al-šbiḥat** rsūmātahum
 Ad: هذوك الرسامون رايعين يوريو الصباح واش رسمو
 Tr: haḍūk al-rsāmūn rāiḥīn iwarīw al-šbāḥ wāš rsmū
 Fr: Ces dessinateurs, ils vont montrer, dans la matinée, ce qu'ils ont dessiné
- P4. Tifinagh: ⵏⵉⵎⵓⵏⵉⵏ ⵏ ⵜⵉⵎⴻⵣⴰⵢⵜ ⵏ ⵜⵉⵎⴻⵣⴰⵢⵜ ⵏ ⵜⵉⵎⴻⵣⴰⵢⵜ ⵏ ⵜⵉⵎⴻⵣⴰⵢⵜ.
 Tr-lat. Ass-nni n tmeɣra, taddart merra **tettwaæred**.
 Ar: نهار العرس الدشرة كامل معروضة :
 Tr: nhār al-ʿrs al-dešra kāmēl **maʿrūda**
 Fr: Le jour de la fête, tout le village était invité.
- P5. Tifinagh: ⵏⵉⵎⵓⵏⵉⵏ ⵏ ⵜⵉⵎⴻⵣⴰⵢⵜ ⵏ ⵜⵉⵎⴻⵣⴰⵢⵜ ⵏ ⵜⵉⵎⴻⵣⴰⵢⵜ ⵏ ⵜⵉⵎⴻⵣⴰⵢⵜ.
 Tr-lat. Asmi d-yuṣal Omer, yufa-d tamsalt **tecbek**
 Ar: يوم عاد عمر وجد القضية مشتبكة
 Tr: iūm ʿad ʿumar waġada al-qaḍia **muštabaka**
 Fr: Le jour où Omar est revenu, il a trouvé l'affaire aggravée.

3.1.3.2 Articles de dictionnaire

À partir d'une source informatisée disponible comme le dictionnaire FEM¹¹ (français-anglais-malais), nous avons construit d'une manière similaire quelques articles de dictionnaire qui seront mis dans la base de données lexicale en JIBIKI. L'exemple suivant

¹¹ <http://www-clips.imag.fr/geta/services/fem/>

est un article de dictionnaire en berbère avec transcription phonétique et sa traduction multicible.

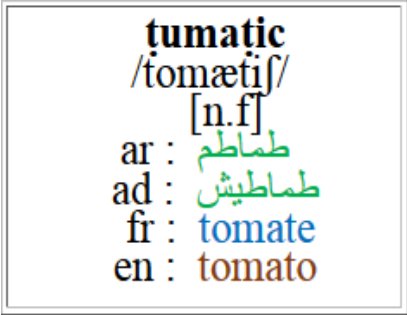


Figure 5 : Article de dictionnaire berbère avec la traduction en plusieurs langues cibles

3.1.3.3 Cognats

Pour les cognats, nous avons créé à la main un tableau en affichage vertical des cognats repérés dans les phrases exemples, avec une coloration syntaxique différente pour chaque langue. Ce tableau représente la ressource.

berbère latin	berbère tifinagh	arabe classique	arabe dialectal	français	Anglais
tilibizyun		التلفزيون	التلفزيون	télévision	television
radio		الراديو	الراديو	radio	radio
tlummun		يلومون	يلومو		
tasebhit		الصبيحة			
tettwaerɛd	ⵜⴰⵝⴰⵎⴰⵏⵜ		معروضة		
tecbek	ⵜⴰⵎⴰⵏⵜ		مشتبكة		

Figure 6 : Capture d’écran d’une partie du tableau des cognats

3.2 Simulation (maquettage)

3.2.1 Objectifs

Les objectifs de cette simulation sont les suivants.

D’abord, il s’agit de nous aider à construire le "cahier des charges" puis les spécifications externes de CESSELIN/ETYMOLO. Cette interface simulée nous permettra de mettre au point ces deux documents (qui feront partie de notre rapport de M2P) avec nos "clients" (LIG-GETALP).

Deuxièmement, cela nous permettra de discuter avec des participants à TALAf en disposant d’une maquette concrète (même si elle n’est pas fonctionnelle), et peut-être de recueillir des idées intéressantes.

Enfin, cela nous permettra de faire une première estimation du travail à faire, et de proposer si possible des façons de le mener à bien faire de façon efficace.

3.2.2 Traitement des cognats

Dans l’interface proposée, nous présentons une phrase en berbère sur la première ligne d’un tableau, et les cognats éventuels sous les mots correspondants dans les autres langues, sur autant de lignes que de langues. Pour l’instant, nous avons créé à la main un tableau de ce type pour chacune de nos phrases exemples.

Voici ce qu’on pourrait voir si on s’intéressait seulement aux cognats, en mode LAE.

Berbère	Aṭas	n	yimawlan	i	yettlummun	arraw-nsen
Arabe					يلومون	
Translittération					īalūmūn	
Arabe dialectal					يلومو	
Translittération					īlūmu	
Français						

Tableau 6 : Tableau de mise en relief des cognats pour la phrase P2

Nous avons construit la première version de ce tableau en même temps que nous avons cherché de bons exemples de phrases à utiliser. Au fur et à mesure de la lecture de quelques pages du corpus de recherche (le journal électronique "La Dépêche de

Kabylie"¹²), nous avons consigné dans un simple tableau Word les cognats que nous trouvions, même si nous ne retenions pas la phrase courante comme phrase exemple. C'est pourquoi nous avons plus de cognats que ceux qui apparaissent dans nos phrases exemples.

3.2.3 Maquette de l'interface complète

La figure suivante présente la maquette que nous avons faite, pour la présentation des cognats seuls. Les cognats en rouge sont de l'arabe dialectal, les cognats en bleu sont du français.



Figure 7 : Capture d’écran de mise en relief des cognats

3.3 Construction de la maquette

3.3.1 Production de pages Web simulant l'interface de CESSELIN/ETYMOLO

Pour faire des maquettes des interfaces qu’on pourrait vouloir fournir, une solution aurait pu être de dessiner ces interfaces. Nous avons préféré modifier le code de la page Web reçue lors de l’utilisation existante (CESSELIN/LECTURE), et remplacer les textes affichés en japonais par les textes correspondants dans les langues visées par notre projet.

Nous avons de plus ajouté quelques éléments d’interface (cases à cocher, etc.) et adapté les textes des boutons. Le résultat n’est qu’une maquette, et si on clique sur les boutons, l’état de la page courante est perdu car elle est remplacée par ce que le serveur JIBIKI répond à l’action du formulaire. Mais cela permet de bien imaginer ce qui pourrait être réalisé. Nous avons travaillé sous FIREFOX DEVELOPER EDITION, sous la direction de Valérie Belynck, qui nous a montré comment faire.

Voici un exemple montrant la phrase (P2) en :

- lecture active,
- avec cognats seuls,
- en lecture active avec les cognats "intégrés".

Berbère	<i>Aṭas</i>	<i>n</i>	<i>yimawlan</i>	<i>i</i>	yettlummun	<i>Arraw-nsen</i>
Arabe classique	الكثير	من	الأولياء	الذين	يلومون	هم-أولاد
Translittération	<i>al-kaṭīr</i>	<i>min</i>	<i>al-awliāʾ</i>	<i>al-laḍī</i>	ialūmun	<i>Awlāda-hum</i>
Arabe dialectal	بزاف	من	لوالدين	اللي	يلومو	هم-ولاد
Translittération	<i>bzāf</i>	<i>men</i>	<i>lwāldīn</i>	<i>lli</i>	ilūmū	<i>wlād-hūm</i>
Français	Plusieur	parmi	parent _{pluriel}	qui	blâmer	enfants-leurs

Tableau 7 : Exemple de la lecture active avec les cognats pour la phrase (P2)

3.3.2 Textes

Les textes que nous avons insérés dans l’interface sont les 5 phrases en berbère que nous avons choisies dans 5 textes du journal électronique "La Dépêche de Kabylie", et deux phrases en tifinagh extraites d’un sujet du bac blanc 2017 (Office national des

¹² <http://www.depechedekabylie.com/>

examens et concours)¹³.

Nous avons créé un fichier source pour chaque phrase (P1, P2..., Pi...), nommé Pi.docx (sauvé aussi en Pi.html), ainsi qu'un fichier Excel pour la lecture active (Pi-LA.xls et Pi-LA.html, un fichier pour la lecture active avec les cognats en couleurs (Pi-cognats.xls et Pi-cognats.html).

On a besoin de générer la forme Pi-cognats car les cognats ne sont pas toujours de bonnes traductions, affichables pour la LA. Par exemple, contaminación (ES) ne veut pas dire la même chose qu'en français (contamination) : le cognat sera "contamination" et l'équivalent sera "pollution".

	A	B	C	D	E	F
1	berbère latin	berbère tfinagh	arabe	arabe dialectal	français	anglais
2	ttlummun		يلومون	يلومو		
3						

Figure 8 : Exemple du fichier P2-COGNATS.XLSX

3.3.3 Cognats

Pour montrer les cognats des 5 phrases que nous avons choisies, nous avons créé à la main 5 tableaux pour les 5 phrases. Nous avons mis dans la première colonne les cognats en berbère. Dans les colonnes suivantes, nous avons mis les langues cibles : arabe classique, arabe dialectal, français et anglais. Pour cela, nous avons utilisé l'éditeur SEAMONKEY COMPOSER. Nous avons mis une coloration syntaxique différente pour chaque langue.

Voici un exemple sur la phrase P2.

berbère latin	berbère tfinagh	arabe classique	arabe dialectal	français	Anglais
ttlummun		يلومون	يلومو		

Tableau 8 : Tableau des cognats pour la phrase P2

3.3.4 Articles de dictionnaire

Nous avons pour l'instant créé à la main les articles de dictionnaire dans le format XML CDM utilisé par JIBIKI, simplement pour l'éditer pour les suivants, nous utilisons l'interface en ligne de PAPILLON/JIBIKI.

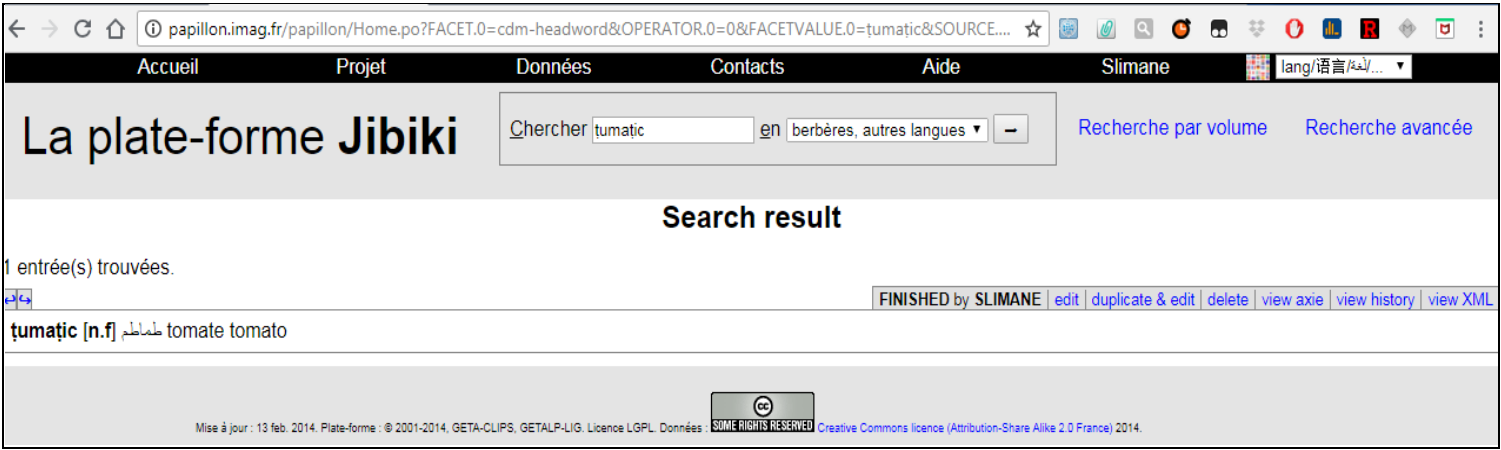


Figure 9 : Capture d'écran d'un article de dictionnaire dans Jibiki

3.4 Exemple de la maquette

Nous avons créé à la main la maquette en utilisant FIREFOX DEVELOPER EDITION pour afficher l'interface JIBIKI/LECTURE pour le japonais-français. Ensuite, Nous avons modifié le code pour afficher une phrase en tfinagh et ajouté des cases à cocher (pour choisir ce qu'on veut seulement voir à l'écran) et nous avons ajouté un article de dictionnaire pour le dernier mot qui est un cognat entre e berbère et l'arabe, plus sa traduction en français et anglais.

¹³ <http://www.exoco-lmd.com/tamazight/sujets-bac-algerie-et-correction-tamazight>


```

style="color: pink; padding-right: 20pt">
tiffinagh</span>
        <input id="writingSelect"
name="écriture" value="tamamrit" type="checkbox"><span
style="color: purple; padding-right: 20pt">
tamamrit</span>
        <input id="translittérationSelect"
name="translittération" value="translittération"
type="checkbox"><span style="color: black; padding-
right: 20pt"> catégorie</span>
        <input id="translationSelect"
name="translation" value="arabe" type="checkbox"><span
style="color: green; padding-right: 20pt"> arabe</span>
        <input id="translationSelect"
name="translation" value="arabe dialectal"
type="checkbox"><span style="color: green; padding-
right: 20pt"> arabe dialectal</span> <br/>
        <input id="translationSelect"
name="translation" value="français"
type="checkbox"><span style="color: blue; padding-
right: 20pt"> français</span>
        <input id="translationSelect"
name="transcription" value="anglais"
type="checkbox"><span style="color: brown; padding-
right: 20pt"> anglais </span>
        <input id="translitterationSelect"
name="translittération" value="translittération"
type="checkbox"><span style="color: red; padding-
right: 20pt"> phonétique arabe</span>
        <input id="translitterationSelect"
name="translittération" value="translittération"
type="checkbox"><span style="color: orange; padding-
right: 20pt"> phonétique arabe dialectal</span>

        <fieldset>
        <legend>Saisissez un texte en
berbère</legend>
        <div style="padding: 1em;">
        <textarea name="text" cols="60" rows="15"
style="font-size: 120%; width: 754px; height: 29px;">
ⵟⵉⴼⵉⵏⴰⵖ ⵏ ⵙⵉⴽⵓⵏ ⵏ ⵉⵎⵓⵔ ⵏ ⵉⵎⵓⵔ ⵏ ⵉⵎⵓⵔ
</textarea>
        <br>
        <select name="langtext"
onchange="showhidetranscription(this.value);">
        <option value="ber"
selected="selected">berbère</option>
        <option value="fra">français</option>
        </select>
        <input name="submit" value="Afficher dans le
texte" type="submit">
        <!--select id="transcriptionSelect"
name="transcription">
        <option value="latin"
selected="selected">latin</option>
        </select-->

        </div>
        </fieldset>
        </form>
</section>

```

```
<style>
</style>
```

•ΘΣ

```

    {V.}
  </span-blue>
  <span-green class="translation" style="color:
green; display: none; margin-top: 70px;">
    عداد، رجع {ف}
  </span-green>
  <span-red class="translation" style="color:
blue; display: none; margin-top: 100px;">
    {V.} revenu, retourné
  </span-red>
  <span-brown class="translation" style="color:
brown; display: none; margin-top: 130px;">
    {V.} returned, came back
  </span-brown>

```

んぞ。O,
<ruby>

 $\Sigma \mathbb{H}_0 - \Lambda$

†.□⊙.И†

```
<span-blue class="translation" style="display:
none;">
</span-blue>
<span-purple class="translation"
style="color: purple; display: inline;">
{Adj.} tecbek
</span-purple>
<span-green class="translation" style="color:
green; display: inline; margin-top:70px">
{مشتبكة، مفقودة}
</span-green>
```

<rbc>

```

    <rb> <strong> т:СЭ:К </strong>
</rb><rp>(</rp>
</rb>
<rtc>
    <rt> <font size="-1"> techbek </font> </rt>
</rtc>
</ruby>
</span>

```