

## Projet Scripta-PSL (2019)

# ANA ŪSĪ ŠAPĀRUM : ÉCRIRE SELON LES RÈGLES EN MÉSOPOTAMIE

PAR M. BÉRANGER (POSTDOC, COLLÈGE DE FRANCE)

C'est la fin du projet ! Grâce au financement octroyé par Scripta-PSL, il est désormais possible d'exporter les milliers de textes lemmatisés dans la base de données Archibab ([www.archibab.fr](http://www.archibab.fr)) pour obtenir des fichiers au format xml enrichis d'une annotation linguistique. Des feuilles de style (xslt) ont été créées pour réorganiser les données d'Archibab et en ajouter de nouvelles.

Index par enrichissement:	Régénérer la liste ci-dessous	Export TXM	Env col.	FTR ligne:	transcription
1: VZ ana à, vers, pour			0	F 1	a-na i-dí-ia-[t]im
1: PH Iddiyatum			0	F 2	qí-bí-ma
2: VV qabúm dire, parler, ordonner			0	F 3	um-ma (d)NU.MUŠ.DA-na-ah-ra-ri
3: VZ umma ainsi (parle)			0	FT 4	i-na (giš)MÁ Ì
3: PH Numušda-nahrari			0	FT 5	ša tu-i
4: VZ ina dans, vers			0	R 6	6 SU KÙ.BABBAR
4: VS eleppum bateau			0	R 7	ma-ki-is (UŠ)
4: VS šamnum huile			0	R 8	wa-aš-še-er
5: VZ ša de			0	RT 9	ITI li-li-ia-tim
5: PH Tu'u			0	RT 10	U. 2.KAM
6: VS šiqlum sicle					
6: VS kaspum argent					
7: VV makâsum percevoir, rassembler					

Vue du texte ARM 13 60 dans Archibab

1. Source : base de données Archibab

2. Étape intermédiaire : feuilles xslt

3. Résultat : fichier xml enrichi d'une annotation linguistique

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-model href="https://raw.githubusercontent.com/MarineBeranger/xml-cuneiform/master/schema" />
<text id="ARM 13 60" genre="lettre" placeComposition="Inconnu" placeDiscovery="Mari"
period="oldbab" king="Zimri-Lim" century="18" year="1776" scribe="TBD"
author="Numušda-nahrari" genderAuthor="TBD" prof="TBD" addressee="Iddiyatum"
genderAddressee="TBD">
<w line="1" lang="akk-x-olbab" frag="a-na" form="a-na" cf="ana" norm="TBD" pos="VZ"
gw="à, vers, pour">
<gdl>
<g value="a" break="no" role="syll" refSyllab="#MesZL839" refUnicode="#U+12000"/>
<g value="na" break="no" role="syll" refSyllab="#MesZL110" refUnicode="#U+1223E"/>
</gdl>
</w>
<w line="1" lang="akk-x-olbab" frag="i-dí-ia-[t]im" form="i-dí-ia-tim" cf="Iddiyatum"
norm="TBD" pos="PH" gw="Iddiyatum">
<gdl>
<g value="i" break="no" role="syll" refSyllab="#MesZL252" refUnicode="#U+1213F"/>
<g value="dí" break="no" role="syll" refSyllab="#MesZL119" refUnicode="#U+12077"/>
<g value="ia" break="no" role="syll" refSyllab="#MesZL260" refUnicode="#U+12140"/>
<g value="tim" break="damaged" role="syll" refSyllab="#MesZL167" refUnicode="#U+12074"/>
</gdl>
</w>
<gap/>
<w line="2" lang="akk-x-olbab" frag="qí-bí-ma" form="qí-bí-ma" cf="qabúm" norm="TBD" pos="V"
gw="dire, parler, ordonner">
<gdl>
<g value="qí" break="no" role="syll" refSyllab="#MesZL737" refUnicode="#U+121A0"/>
<g value="bí" break="no" role="syll" refSyllab="#MesZL313" refUnicode="#U+12248"/>
<g value="ma" break="no" role="syll" refSyllab="#MesZL552" refUnicode="#U+12220"/>
</gdl>
</w>
</text>
```

Vue du texte ARM 13 60 dans Oxygen XML Editor

### MÉTADONNÉES, LEMMATISATION ET PART-OF-SPEECH TAGGING EXTRAITS D'ARCHIBAB :

- métadonnées (<text>) : référence-clef du texte, genre, lieu de rédaction, lieu de découverte, date, nom de l'auteur du texte, etc.
- lemmatisation et part-of-speech tagging (<w>) : graphie du mot, lemme, catégorie grammaticale et traduction.

### NOUVEAUTÉS PAR RAPPORT À ARCHIBAB :

- annotation au niveau du signe cunéiforme (<g>) : lecture de chaque signe en contexte, visibilité du signe sur la tablette d'argile (visible, à demi visible, disparu), rôle (logogramme, syllabogramme, déterminatif, etc.), référence dans le syllabaire de Borger et référence Unicode.

## ET APRÈS ?

— Un premier lot de textes annotés sera mis en ligne sur la plateforme Ortolang (<https://www.ortolang.fr>) dans un proche futur.

— Les fichiers produits dans le cadre de ce projet peuvent par exemple être utilisés pour comparer le vocabulaire des textes administratifs / lettres / textes juridiques, connaître la manière dont un mot était orthographié dans différentes villes / à différentes époques, ou encore connaître les différentes lectures d'un même signe cunéiforme. Pour un exemple de recherche possible, voir : M. Béranger, « Glimpses of the Old Babylonian Syllabary. Followed by Some Considerations on Regional Variations and Training in Letter-Writing », in J. Klinger & S. Fischer (éd.s), BBVO 28, 2019, p. 17-38.

— Le logiciel «TXM», développé à l'ENS de Lyon, est recommandé pour travailler sur ces fichiers, car il est compatible avec le xml, est gratuit et offre de nombreux outils statistiques (<http://textometrie.ens-lyon.fr>) — mais d'autres logiciels peuvent bien sûr être utilisés.