

**Traitement informatique en assyriologie :
un corpus des textes cunéiformes
d'époque babylonienne ancienne**

par

Philippe TALON

Université Libre de Bruxelles - BELGIQUE

Le but de cette communication est de présenter les réalisations et les projets du G.I.T.A. (Groupe Facultaire d'Informatique et de Traitement Automatique, Université Libre de Bruxelles) dans le domaine de l'assyriologie. Et il faut bien reconnaître que les progrès de l'informatique n'ont guère touché cette discipline.

Rappelons que l'assyriologie a pour objet l'étude philologique et historique des documents rédigés en caractères cunéiformes, découverts dans tout le Proche-Orient, de l'Égypte à l'Anatolie et de la Méditerranée à l'Iran.

L'informatique a fait son entrée dans ce domaine de manière très discrète. Quelques travaux ont été publiés, à caractère surtout bibliographique ⁽¹⁾. Un ouvrage récent fait exception par son ampleur : I. J. GELB, *Computer-aided Analysis of Amorite*, *Assyriological studies* n. 21, Chicago, 1980. D'autres projets commencent à s'annoncer dont les résultats n'ont pas encore été rendus publics ⁽²⁾.

Grâce aux moyens mis à ma disposition par le G.I.T.A. à l'U.L.B., et grâce aux programmes écrits et mis au point par R. Patesson, j'ai pu mettre en chantier un corpus de textes akkadiens directement utilisables par l'ordinateur de l'U.L.B. ⁽³⁾.

1.1. Depuis plusieurs années déjà, le G.I.T.A., sous la direction de J.-H. Michel, s'est doté de programmes de concordance, d'index et de tri. Ceux-ci ont été réalisés dans le but de pouvoir s'adapter le plus aisément possible à des textes rédigés en des langues différentes, anciennes ou modernes.

1.2. La souplesse d'utilisation de ces programmes s'est trouvée démontrée encore lorsque, dans le cadre de ma thèse de doctorat, j'ai tenté de les appliquer à ce domaine particulier de la philologie qu'est l'assyriologie ⁽⁴⁾.

2.1. Le caractère un peu particulier de l'application du traitement automatique à l'assyriologie tient en majeure partie au support d'écriture qui nous a transmis les langues anciennes du Proche-Orient : l'écriture cunéiforme.

2.2. Celle-ci s'est développée rapidement à partir de la fin du 4^{ème} millénaire av. J.-C., pour rester en usage, à travers une longue évolution, jusqu'au premier siècle de notre ère. Elle nous a livré des langues diverses, appartenant à des familles linguistiques bien connues, telles le sémitique ou l'indo-européen, ou à des groupes mal définis que l'on classe par commodité, sous le vocable "asianique", tels le sumérien, le hurrite ou l'élamite.

2.3. La langue la plus anciennement attestée est le sumérien, et, sans entrer dans les détails, il apparaît que l'écriture cunéiforme a été mise au point par les Sumériens. Elle fut rapidement transmise à la langue sémitique qui supplanta le sumérien en Mésopotamie et qui devait devenir la langue véhiculaire et diplomatique de tout le Proche-Orient à partir du 2^{ème} millénaire : l'akkadien. L'akkadien est, avec l'éblaïte récemment découvert en Syrie, la langue sémitique la plus ancienne connue.

2.4. L'écriture cunéiforme se compose de deux éléments majeurs pour former un système logographique ou morphémo-phonographique. Logographique ou morphémographique, car elle dispose de signes ou graphèmes qui représentent des unités linguistiques signifiantes (morphèmes)⁽⁵⁾; phonographique, car elle dispose de graphèmes qui représentent des unités linguistiques non-signifiantes, des phonèmes et des groupes de phonèmes sous la forme de syllabes ouvertes et fermées⁽⁶⁾.

2.5. Ces graphèmes remontent pour la plupart à l'état le plus ancien de l'écriture et dérivent du sumérien pour leur état phonologique. Ils furent lentement modifiés et aménagés pour rendre compte des faits linguistiques propres à l'akkadien⁽⁷⁾. Ainsi, à la différence de la plupart des systèmes d'écriture appliqués aux langues sémitiques, le cunéiforme note soigneusement la vocalisation et non pas uniquement le "squelette" consonantique des mots, comme c'est le cas en araméen, en hébreu et en arabe (compte non tenu des signes diacritiques spéciaux vocalisant la Bible ou le Coran, surtout, et qui sont une adjonction indépendante du système d'écriture lui-même). Ceci rend l'étude de l'akkadien particulièrement importante dans le cadre de l'évolution des langues sémitiques.

2.6. Une transcription de texte cunéiforme se présente donc pour nous comme une juxtaposition de graphèmes syllabiques, avec une présence numériquement variable de logogrammes sumériens. Exemple : [ARM X.26 : 5 - 7] *tu-i-mi ú-[ta-a]-li-id 1 DUMU ù MI.DUMU-tam be-lí lu-ú ha-di* [=tu' imi utallid isten maram u martam; beli lu hadi'], "je [viens d'ac]oucher de jumeaux, un garçon et une fille. Que mon seigneur se réjouisse !" (les logogrammes sont transcrits en majuscules).

3.1. La documentation qui nous a été transmise par le biais de l'écriture cunéiforme est extrêmement vaste. Le nombre total de documents retrouvés depuis le siècle dernier n'est pas connu avec précision, mais on a pu avancer un chiffre de l'ordre de 200.000 lignes⁽⁸⁾ (ce chiffre ne tient pas compte du fait que certains textes sont connus par de nombreux duplicats, il est donc sous-estimé). Celui-ci est en perpétuelle augmentation par d'incessantes découvertes de textes nouveaux sur les sites archéologiques du Proche-Orient.

3.2. A l'intérieur de cette gigantesque documentation, des spécialisations se sont créées, selon la langue utilisée, l'époque ou le type de texte : archives administratives, économiques, lettres, textes religieux, mythologiques, historiques, mathématiques, astronomiques, etc.

3.3. Le corpus que j'ai choisi d'étudier et d'enregistrer sur ordinateur est formé par un groupe cohérent de documents rédigés en akkadien, découverts sur le site de l'ancienne ville de Mari, aujourd'hui Tell Hariri à la frontière syro-irakienne, et remontant aux années 1800-1750 avant notre ère. Ce sont essentiellement des archives administratives et économiques et des lettres privées et diplomatiques qui se répartissent sur deux règnes d'une durée d'environ 40 ans. Une partie seulement de ces archives a été publiée, mais leur édition est en cours. Nous disposons actuellement d'environ 3.500 tablettes, soit $\pm 1/4$ du total⁽⁹⁾.

3.4. Afin de pouvoir étudier les différents aspects de ces documents, ils ont été enregistrés sur l'ordinateur de l'U.L.B. avec une série de codes permettant d'isoler les logogrammes sumériens, les

anthroponymes, toponymes, théonymes, etc, tout en conservant au maximum le format et la lisibilité des textes eux-mêmes (10).

3.5. Actuellement, j'ai pu enregistrer la totalité des archives administratives auxquelles j'ai ajouté les textes de même nature dont la publication m'incombe. Le travail se poursuit maintenant avec les documents épistolaires. Il est vraisemblable que l'ensemble des tablettes publiées sera disponible en fichier dans le courant de l'année 1982. Ce corpus, qui contient déjà plus de 50.000 mots, devrait croître rapidement par l'adjonction de textes en cours de publication (11).

4.1. Les réalisations qui ont déjà pu être obtenues sur le corpus existant sont encourageantes. Les programmes d'index et de concordance ont donné des résultats satisfaisants malgré les difficultés dues aux particularités du système d'écriture et de transcription. Les possibilités d'index inverse se sont révélées particulièrement intéressantes pour la restitution des passages lacuneux. De même, les concordances ont mis en lumière des formules et des convergences qui permettent aussi de restaurer les lignes en mauvais état de conservation.

4.2. De plus, et grâce à l'aspect syllabique de la transcription, on a pu obtenir des listes de graphèmes, syllabiques et morphémographiques, accompagnées de statistiques d'utilisation. C'est là un aspect quelque peu inattendu du programme d'index, qui a tenu simplement à quelques changements dans la déclaration des séparateurs de mots. On peut y trouver matière à l'étude des syllabaires locaux, dont on sait qu'ils ont varié considérablement d'une époque à l'autre, ou même géographiquement d'un royaume à l'autre.

5.1. La structure de l'écriture et de la langue akkadienne se prête mal à des tentatives de lemmatisation automatique. Héritée du sumérien, l'écriture cunéiforme se révèle mal adaptée à rendre certaines particularités phonologiques des langues sémitiques (notamment les consonnes emphatiques). D'autre part, l'akkadien a modifié quelque peu son système phonétique sous l'influence du sumérien. Ces caractéristiques rendent la recherche automatique des racines consonantiques extrêmement difficile; celles-ci sont souvent cachées ou modifiées par assimilation et dissimilation ou encore sous-entendues et illustrées dans la graphie par une simple rupture de séquence syllabique (12). Malheureusement, ces ruptures de séquence ne constituent pas une preuve en soi et peuvent recevoir d'autres explications, selon les cas.

5.2. L'acquisition directe et automatique d'un dictionnaire lemmatisé se révèle donc difficile. Il faudra passer par des fichiers intermédiaires où il sera possible d'injecter les définitions grammaticales et les traductions, pour obtenir ensuite, par tris successifs, une liste exhaustive du vocabulaire. Ce projet est à l'étude.

5.3. Il est à espérer qu'il sera réalisable, car ses applications laissent entrevoir des études grammaticales très intéressantes. L'isolation de formes, verbales ou autres, avec leur contexte, par tris divers, donnerait au grammairien un instrument de travail incomparable et surtout dynamique.

5.4. L'impératif numéro un reste dans l'immédiat, l'achèvement de l'enregistrement du corpus. Les concordances et les index sont au point et permettent déjà de fructueuses études partielles, ainsi que la restitution de certains passages abîmés et l'obtention de listes de signes qui restent très précieuses.

5.5. Il faut espérer que ce corpus pourra ensuite s'élargir pour recevoir les documents contemporains de la Première Dynastie de Babylone. Nous comptons en Belgique des spécialistes de cette période et cet élargissement nécessitera certainement une collaboration interuniversitaire. Je souhaite vivement que cette collaboration se révèle possible; elle sera certainement fructueuse, comme l'a été et l'est encore la collaboration entre les moyens modernes de l'informatique et une science philologique traditionnelle, l'assyriologie.

NOTES

- (1) Par exemple : J.-G. HEINTZ, *Index documentaire des textes de Mari, Archives Royales de Mari*, vol. XVII/1, Paris, 1975. Les programmes utilisés dans cet ouvrage ont été mis au point par Mme C. CROZIER-BRELOT (cf ARMT XVII/1, p. 10*); on lira à ce propos la communication de Mme CROZIER-BRELOT à ce congrès.
- (2) Notamment les projets de G. BUCCELLATI à Los Angeles.
- (3) Cf. Ph. TALON, *Assyriologie et Informatique, Akkadica* 23, Bruxelles, 1981, p. 23-36; R. PATESSON, *Le traitement informatique des données en sciences humaines : quelques principes*, ibidem, p. 37-44.
- (4) *Textes administratifs des salles Y et Z du Palais de Mari*, 3 volumes, thèse ULB 1981.
- (5) Ex. : LUGAL = roi, LU = homme, DU = créer, etc.
- (6) Du type CV, vc, CVC (les signes de type CVCV(C) sont très rares, surtout à l'époque qui nous intéresse ici).
- (7) Ex. : KA sumérien "bouche" → /ka/, akkadien pûm → /pi4/; UD/UTU sumérien "soleil" → /ud/, /ut/, /ut/, /tam/, etc.
- (8) A.L. OPPENHEIM, *Ancient Mesopotamia. Portrait of a Dead Civilization*, Chicago, 1977², p. 18.
- (9) *Archives Royales de Mari* (2 séries : textes en transcription et copies cunéiformes), Paris, 1950- (19 volumes parus).
- (10) Codes : noms propres de personnes (=), logogrammes (!), toponymes (+), noms divins (#, ^), "aleph" ('), § (C), † (X), § (\$).
- (11) 53198 mots au 1.12.81, soit ARM IV, V, VII, IX, XIII, XVIII et XXIII (à paraître) (ARM IX n'a pas été enregistré au complet pour des raisons particulières); les volumes I, II, X et XIV sont en cours d'enregistrement. Cette documentation est répartie en différents fichiers afin de pouvoir être groupée, selon les besoins, en corpus thématiques : lettres de telle époque, archives économiques, etc.
- (12) ex. : ARM I.89:5 šu-re-eš-ši < *šuwri'am +ši "fais-la moi amener" (impératif S 2ème pers. sg ventif, avec suffixe 3ème pers. fém. sg), tiré du verbe *warûm* (racine wrû).