DiCoNab, the Digital Corpus of Nabataean and Developing Arabic Inscriptions
Laïla Nehmé, Jérôme Norris

To cite this version:
Laïla Nehmé, Jérôme Norris. DiCoNab, the Digital Corpus of Nabataean and Developing Arabic Inscriptions. 2023. halshs-04336551

HAL Id: halshs-04336551
https://shs.hal.science/halshs-04336551
Preprint submitted on 18 Dec 2023

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.
DiCoNab, the Digital Corpus of Nabataean and Developing Arabic Inscriptions

By Laïla Nehmé (CNRS–UMR 8167) and Jérôme Norris (EA 1132 HISCANT–MA, Nancy)

This is the provisional version (before page set up and proofs) of an article to be published (in 2024?) in the Festschrift in honour of D.F. Graf, professor at Miami.

It is a pleasure for us to contribute to this *Festschrift* prepared in honour of David Graf, whom Laila first met in 1992 in Jordan, during one of the first conferences she attended as a graduate student. Since then, their paths crossed many times, either in Jordan, in Paris, or in other places, mainly on the occasion of other conferences. David Graf was always very friendly and willing to share his knowledge and discuss ideas. Growing older, their relationship became that of two colleagues who always remained in touch, exchanged views and asked for each other’s opinion on specific subjects. Presenting DiCoNab, a tool that took far too long to develop, is the best way for us to show him that Nabataean studies, which are dear to his heart, continue to progress, with the decisive implication of younger scholars such as J. Norris. These studies will hopefully gain more efficiency and more popularity among the newer generation by making the documents accessible online and by providing, through computer generated maps and indexes, renewed research opportunities.

After the success of two online databases devoted to Ancient South Arabian and Ancient North Arabian inscriptions entitled the “Corpus of South Arabian Inscriptions” (CSAI)\(^1\) and the “Online Corpus of the Inscriptions of Ancient North Arabia” (OCIANA)\(^2\) respectively, it was high time to develop a database for both the Nabataean and the “Developing Arabic” inscriptions (see below). Besides, the maturity of open source database systems such as MySQL and PostgreSQL, which have proved their reliability over the last twenty years or so, convinced the authors that these highly interoperable systems were indeed the best suited.

As is very well known, Nabataean (or Nabataean Aramaic) is the name given to the language and script used by the Nabataeans, an Arab tribe the presence of which is attested in southern Jordan, in the region of Petra, from the late fourth century BCE onwards. The members of this tribe established a kingdom which, at its greatest extent, reached Damascus in the north and the Hijāz in the south. It remained independent until 106 CE, when the Romans annexed it and reduced it into the Province of Arabia. The Nabataeans produced thousands of inscriptions, the vast majority of which are graffiti carved on stone. Only a little more than twenty Nabataean papyri have survived (Healey 2004). The Nabataeans also minted vast numbers of coins with legends in the Nabataean script (Meshorer 1975 and Barkay 2019).

As for Developing Arabic, it is the name given to two categories of scripts formerly labelled “Nabataeo-Arabic” and “pre-Islamic Arabic” or “Palaeo-Arabic”, found in some of the regions where the Nabataean script continued to be used after 106 CE. “Nabataeo-Arabic” refers to the script which is clearly transitional between Nabataean and Arabic (fig. 1). As such, it contains letter shapes which are still recognisable as Nabataean or closer to Nabataean than

---

1. [http://dasi.cnr.it/index.php?id=42&prjld=1&corld=0&colld=0](http://dasi.cnr.it/index.php?id=42&prjld=1&corld=0&colld=0)
2. [http://krcfm.orient.ox.ac.uk/fmi/webd/ociana](http://krcfm.orient.ox.ac.uk/fmi/webd/ociana)
they are to Arabic. It is the predecessor of “pre-Islamic Arabic” or “Palaeo-Arabic”, most letter shapes of which are already recognisable as Arabic (fig. 2).

Since the distinction between Nabataeao-Arabic and Palaeo-Arabic can be difficult because it is not based on a list of objective criteria, it has been decided to group the two under a single umbrella and make the distinction between them only when there is no ambiguity.

The Developing Arabic inscriptions are dated to the interval between the last quarter of the 3rd and the early 7th century CE. Those which are written specifically in Nabataeao-Arabic are usually not later than the mid-fifth century CE. The number of Developing Arabic inscriptions does not, at present, exceed 250. The majority have been found in what is now Saudi Arabia, with a few examples in Syria and Jordan. Altogether, the countries in which inscriptions written in the Nabataean, Nabataeao-Arabic or Palaeo-Arabic scripts were found are Syria, Jordan, Saudi Arabia, Yemen, Egypt, and Israel. In a much lower proportion, and only in the Nabataean script, inscriptions were found in other countries, namely Greece, Italy, Lebanon, and Turkey.

The DiCoNab project was launched in september 2022 and is based at the UMR 8167, the Orient & the Mediterranean laboratory of the French National Centre for Scientific Research (CNRS, https://www.orient-mediterranee.com/). It is undertaken under the aegis of the Académie des Inscriptions et Belles-Lettres, in the framework of the Corpus inscriptionum semiticarum (https://aibl.fr/travaux/le-cabinet-du-corpus-inscriptionum-semiticarum/). The academic director is Laïla Nehmé (CNRS), with the collaboration of Michael C.A. Macdonald (University of Oxford) and Jérôme Norris (University of Lorraine). It is hosted by Huma-Num (https://www.huma-num.fr/), a French infrastructure which provides support to research projects in the field of digital humanities.

The initial development of the database was made thanks to the support given through Benjamin Suchard’s research project on “Nabataean as a spoken language” (2019–2022, VI.Veni.191T.023, Dutch Research Council).³ In its initial stages, it was also helped by the

³ The project’s director is extremely grateful to B. Suchard for his immediate and unreserved positive answer to our request for support in order to set up the database.
Maparabia project\(^4\) as well as by a remainder of the sponsorship given between 2019 and 2023 by the Total Foundation to the Madain Salih Archaeological Project.

**Users**

DiCoNab is of course addressed to epigraphists, archaeologists, and historians. It will make it easy to find much of the information they may need, from dated inscriptions to the mention of any word(s), names, etc., in the texts. This is normal and will not be more detailed. The aim is also to provide either Master/PhD students or academics in charge of the study or publication of corpus(es) of texts the possibility to enter in the database the texts they work,\(^5\) and thus benefit from its processing power to build indexes of all sorts and analyse the texts’ contents from various perspectives. Surely, the internet pages of the texts entered as part of such studies or publications should not be visited by all users indiscriminately but only by those who are entering them and by the community of specialists. By putting the material on an online device, students and academics would thus benefit from the help of colleagues but the material would not be accessible by all users until the author decides to make it public. This is simply done clicking on either “Public” or “Private” in a dedicated field.

**Technical aspects**

*The database was developed by Matteo Gallo, a freelance developer from Pisa who has gained, with CSAI and other projects, a real and extremely valuable experience in academic databases, including in the field of Semitic studies. The database was developed using MySQL, an open source database management system. Each text is provided with a unique Uniform Resource Locator (URL) which allows to refer to it easily (e.g. [https://diconab.hum-num.fr/inscriptions/14](https://diconab.hum-num.fr/inscriptions/14)).

**Structure**

The structure of DiCoNab is that of relational database. It includes three main enterable related tables: Sites, Inscriptions, and References (publications/bibliographical references). These are related either by 1 to N or by N to N links.\(^6\) Thus, Sites are related to Inscriptions by a 1 to N link because one site can contain several inscriptions whereas one inscription cannot belong to more than one site. On the other hand, Inscriptions are related to References (publications) by an N to N link because one inscription can appear in one or more publication(s) and vice versa, one publication can contain one or more inscription(s). The link between the two is made through a fourth enterable table called “Editions and commentaries” in which the data relative to a specific inscription in a specific publication is entered.\(^7\) This allows to build automatic indexes of references, as are often provided at the end of proper academic publications of epigraphic material.

In addition, three always accessible utilities offer the significant advantage of being modifiable by the administrators of the database or by the standard users upon simple request to the

---


\(^5\) All they need is to be given access by DiCoNab’s administrator(s) through a login and a password which makes them “Editors” of the database rather than only “Standard users”.

\(^6\) N represents an unknown and unlimited number of records.

\(^7\) Page, copy, photograph, squeeze, reading and translation categories.
administrators, without—and that is important—the help of a computer engineer. These are the List of values, the Word tags, and the Media resources.

<table>
<thead>
<tr>
<th>Word list - JSNab 16</th>
<th>Nb of occurrences in JSNab 16</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I – Preposition</td>
<td>8</td>
</tr>
<tr>
<td>2. P – Negative particle</td>
<td>1</td>
</tr>
<tr>
<td>3. mwrb – Subst. sg. suffixed</td>
<td>1</td>
</tr>
<tr>
<td>4. mstb – Subst. sg. construct</td>
<td>1</td>
</tr>
<tr>
<td>5. mn – Preposition</td>
<td>8</td>
</tr>
<tr>
<td>6. mnrw – Divine name</td>
<td>2</td>
</tr>
<tr>
<td>7. nbtw – Non gentilic group name</td>
<td>1</td>
</tr>
<tr>
<td>8. npst – Subst. sg. suffixed</td>
<td>1</td>
</tr>
<tr>
<td>9. styn – Subst. pl. absolute</td>
<td>1</td>
</tr>
<tr>
<td>10. ‘bd – Suffix conjugation p.t. (G-stem)</td>
<td>1</td>
</tr>
<tr>
<td>11. ‘bdw – Suffix conjugation p.t. (G-stem)</td>
<td>1</td>
</tr>
<tr>
<td>12. ‘bd‘bdt – Man’s personal name</td>
<td>1</td>
</tr>
<tr>
<td>13. ‘y – Adjective sg.</td>
<td>1</td>
</tr>
<tr>
<td>14. ‘p – Adverb</td>
<td>1</td>
</tr>
<tr>
<td>15. (‘)m – Subst. sg. suffixed</td>
<td>2</td>
</tr>
<tr>
<td>16. ‘mn(w) – Place name</td>
<td>1</td>
</tr>
<tr>
<td>17. p – Preposition</td>
<td>1</td>
</tr>
<tr>
<td>18. p – Conjunction</td>
<td>1</td>
</tr>
<tr>
<td>19. qym – Passive participle sg.</td>
<td>1</td>
</tr>
<tr>
<td>20. qyš – Subst. sg. suffixed</td>
<td>1</td>
</tr>
<tr>
<td>21. qns – Subst. sg. absolute</td>
<td>1</td>
</tr>
<tr>
<td>22. ṛhm – Active participle sg.</td>
<td>1</td>
</tr>
<tr>
<td>23. šl – Subst. sg. absolute</td>
<td>1</td>
</tr>
<tr>
<td>24. šmdyn – Subst. pl. absolute</td>
<td>1</td>
</tr>
<tr>
<td>25. šnt – Subst. sg. absolute</td>
<td>1</td>
</tr>
<tr>
<td>26. ts – Number in letters</td>
<td>1</td>
</tr>
</tbody>
</table>

Fig. 3. Word list of JSNab 16 generated automatically.
The List of values utility concerns fields—such as Country, Script, Writing technique and several others—that are enterable through predefined values. For example, the field “Associated motif(s)” in the Inscriptions’ table contains the following values: Betyl, Cavalryman, Dromedary, Eagle, Equid, etc.”. Like all lists of values, they allow data to be entered in a fully consistent way.

Word tags are the tags which can—and are highly recommended to—be entered at the end of each word in the transliteration of an inscription. Each word tag refers to a particular word category (substantives, verbs, proper nouns, particles, etc., see the list of word tags in Appendix 1). The Word tags are also mofiable by the administrators and since some tag values are not used, a few tags can be added if needed.

The tagged version of a transliterated text (in Unicode)\(^8\), here JSNab\(^9\) 16, looks as follows, where the tags always come immediately after the sign “*”:\(^10\)

\[
dnh*66 kpr*13 dy*64 'bdw*22 kmkm*82 brt*12 w'lt*82 brt*12 ḥrmw*81_ w*72 klybt*82 brt*14 -h*62 l*71 npš*14 -hm*63 w*72 ḥr*14 -hm*63 b*71 yrh*12 tbt*91 Šnt*11_ tš*94 l*71 ḥrtt*84 mlk*12 nbtw*88 ṭhm*43 {jm}n*14 -h*62 w*72 yl*9*32 dwšr*87 w*72 mwtn*14 -h*62 w*72 'lt*87 mn*71 mn(w)*89 w*72 mnntw*87 w*72 qyš*14 -h*62 mn*65 yzbn*32_kpr*13 dnh*66 w*72 mn*65 yzbn*32 w*72 yrh(n)*32 w*72 (yn)tn*32 yt*78 -h*62 w*72 ynpq*32 mn*71 -h*62 gt*11 'w*72 šlw*11 'w*72 mn*65 yqbr*32 b*71 -h*62 'yr*51 kmkm*82 w*72 brt*14 -h*62_ w*72 ḥr*14 -hm*63 w*72 mn*65 dy*64 l*74 y'bd*32 kdy*72 l*77 ktyb*45 p*72 yty*77 m*71 -h*62_ l*71 dwšr*87 w*72 hblw*87 w*72 l*71 mnntw*87 šmdyn*15 5*93 w*72 l*71 pkl*13 qns*11 sl'yn*15 'lp*94 ḥrt*51 bl'd*71 mn*65 dy*64 ynpq*32 b*71 yd*14 -h*62 ktb*11 mn*71 yd*12_kmkm*82 w*72 klybt*82 brt*14 -h*62 b*71 kpr*13 hw*66 p*72 qym*45 ktb*13 hw*66 whb'lhy*81 br*0 'bd'bdt*81_ 'bd*22
\]

The non tagged version of the text is of course created automatically and looks as follows:

\[
dnh kpr' dy 'bdw kmkm brt w lt bt ḥrmw w klybt brth lnpshm w ḥrmw byrḥ tbt šnt tš' ḥrtt mlk nbtw ṭhm '{jm}n w yl'n dwšr' mw mntt w yqšš mn yzbn kpr' dnh' w mn yzbn' w yrh(n) w (yn)tn yṭh' w ynpq mnh gt 'w šlw' w mn yqbr bh 'yr kmkm w brth w ḥrmw w mn dy l' y'bd kdy l' ktyb p' yty' mh lwšr' h hblw w mnntw šmdyn 5 w pkl' qns sl'yn 'lp ḥrt' bl'd mn dy ynpq bydh ktb mn yd kmkm' w klybt brth bknpr' hw pùmm ktb' hw whb'lhy br 'bd'bdt 'bd.
\]

Tagging a text immediately provides a list of all the words and nouns that appear in it (fig. 3). Finally, two additional non enterable tables (i.e. their contents is updated automatically and cannot be modified by the users) complete the structure of the database: Index and Attestations. The difference between the two is simple but very important.

The Index contains one single entry for each different word or noun that appears in the inscriptions. For example, the conjunction ‘w, “or”, will appear only once in the Index even if

\[^8\] Special user-friendly keyboards for Azerty and Qwerty and for both Apple and PC computers have been developed so that entering letters with diacritical marks (’,’, h, s, ñ, etc.) is both quick and easy.

\[^9\] Jaussen and Savignac 1909–1922, Nabataean inscription no. 16.

\[^10\] Note that the underscore “_” marks a change of line and that enclitic personal pronouns are preceded by “space+hyphen” so that they can be tagged and indexed separately from the word to which they are attached. Thus, npš -hm line 2 is tagged npš*14 -hm*63 because npš here is a substantive singular suffixed and -hm is an enclitic personal pronoun plural.
it attested a number of times in the Nabataean inscriptions. This table is thus more or less the equivalent of the entries of a potential dictionary. A simple click on the word then lists all the attestations/occurrences in each text, in the order of appearance (fig. 4). One of the great assets of the system is that it takes into account the presence of editorial signs such as the curly brackets { } for uncertain letter(s) and the square brackets [ ] for restored letters. Thus, if a word is attested only once and if this single attestation contains either uncertain or restored letters, it appears in red in the Index (fig. 5) in order to draw the user’s attention to the fact that its existence in the corpus is not absolutely sure. In the example given fig. 5, the letters (yn)tn—of the verb ntn in the Prefix conjugation pʿl (G-stem), tag *32—are uncertain and the fact that yntn appears in red in the list shows that this derived form of ntn is not surely attested in Nabataean.11

All the attestations/occurrences of the words are entered automatically in the other non-enterable table, Attestations. An asset of this table is the fact that the words always appear in their context, and whatever comes after the word is given in the Aramaic alphabetical order (fig. 6). This is particularly useful for personal names because it allows to find individuals who bear the same name and whose father’s name (or even grandfather’s name, etc.) is also the same, thus providing the means to trace the same individuals, particularly between epigraphic sites on the trade routes and between the major Nabataean cities and these trade routes.

11 This is of course only an example since (yn)tn is in fact widely attested in Nabataean!
In conclusion, lists of words and nouns, given with their grammatical or noun category and in their context, provide continuously updated versions of the Nabataean dictionaries and corpora of personal names\textsuperscript{12} and may form the basis of a new grammar of Nabataean, which would add a lot of material and examples to the one by J. Cantineau published in 1930–1932.

**Other assets of DiCoNab**

**Sites**

About 520 sites where Nabataean inscriptions were found are already entered in the database, along with their geographical coordinates. The overwhelming majority of these sites were listed by J. Norris who has been collecting, for years, sites with epigraphic material of various sorts in the Middle East. Only those sites which have yielded inscriptions that are already published are made available to all users the database. The rest will be accessible only to the editors who have received a login and a password.

**Example of information available for each site:**

<table>
<thead>
<tr>
<th>Present name</th>
<th>Ancient name</th>
<th>Country</th>
<th>Region</th>
<th>Accuracy of coordinates</th>
<th>Origin of coordinates</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buṣrā</td>
<td>Bostra</td>
<td>Syria</td>
<td>Hawrān</td>
<td>Certain</td>
<td>Pointed on Google Earth</td>
<td>32.51947</td>
<td>36.48257</td>
</tr>
<tr>
<td>Dūmat al-Jandal</td>
<td>Dūmat (Adumatu)</td>
<td>Saudi Arabia</td>
<td>North-West Arabia</td>
<td>Certain</td>
<td>Pointed on Google Earth</td>
<td>29.8113</td>
<td>39.86904</td>
</tr>
<tr>
<td>Wādī Ramm, Temple of Allāt</td>
<td>Iram</td>
<td>Jordan</td>
<td>Northern Hismā</td>
<td>Certain</td>
<td>Pointed on Google Earth</td>
<td>29.57784</td>
<td>35.41457</td>
</tr>
</tbody>
</table>

The availability of coordinates will allow the production of distribution maps, especially since DiCoNab allows exports to be made towards the maps of the atlas of the MAPARABIA project, directed by J. Schiettercatte (http://ancient-arabia.aidns-ag3.eu/atlas).

**Bibliographical references**

All the bibliographical references known and checked by L. Nehmé which concern Nabataean and Nabataean-Arabic epigraphy, Nabataean papyrology, numismatics, onomastics, as well as general works on the Nabataeans have also already been entered in the database with a siglum and a short title, using a Zotero export format designed especially for DiCoNab:\textsuperscript{13}

**Example of information available for each reference:**

<table>
<thead>
<tr>
<th>Siglum</th>
<th>Short title</th>
<th>Complete reference</th>
</tr>
</thead>
</table>

\textsuperscript{12} Thus superseding those published by W.J. Jobling, S. al-Theeb, and A. Negev.

\textsuperscript{13} 875 references are available to all users.


**Media resources**

Media resources are entered in a separate table and can be called whenever needed. This is particularly useful if a photograph contains more than one inscription and needs to be referred to for each one of them so that the photograph is uploaded only once. The media accepted are images, whether photographs or facsimiles, but also 3D models, videos, PDF documents, audios and fonts, the only limit being the space it occupies on the Huma-Num server.

It is hoped that DiCoNab will meet the needs of all those who are interested in Nabataean studies and in the development of the Nabataean to the Arabic script. Taking on the legacy of scholars such as Jean Starcky and Joseph Milik, who started to write in Latin and by hand the continuation of volume II of the *Corpus Inscriptionum Semiticarum*, needed at least an ambitious project such as DiCoNab, for the elaboration of which we have put a lot of the experience we gained in databases of Semitic inscriptions over almost thirty years.

**References**


Appendix 1: Word tags

*0 Not to be indexed

[All substantives start with 1]
*10 Substantive unspecified
*11 Subst. sg. absolute
*12 Subst. sg. construct
*13 Subst. sg. emphatic
*14 Subst. sg. suffixed
*15 Subst. pl. absolute
*16 Subst. pl. construct
*17 Subst. pl. emphatic
*18 Subst. pl. suffixed
*19 Subst. dual

[All suffix conjugation start with 2]
*20 Suffix conjugation unspecified
*21 Suffix conjugation p’l (G-stem or D-stem)
*22 Suffix conjugation p’l (G-stem)
*23 Suffix conjugation p’l (D-stem)
*24 Suffix conjugation hp’l (C-stem)
*25 Suffix conjugation ’p’l (C-stem)
*26 Suffix conjugation šp’l (C-stem)
*27 Suffix conjugation htp’l (Gt-stem or Dt-stem)
*28 Suffix conjugation ’tp’l (Gt-stem or Dt-stem)
*29 Specifically Arabic verbal forms (tafa’ala, infa’ala, etc.) suffix conjugation

[All prefix conjugation start with “2”]
*30 Prefix conjugation unspecified
*31 Prefix conjugation p’l (G-stem or D-stem)
*32 Prefix conjugation p’l (G-stem)
*33 Prefix conjugation p’l (D-stem)
*34 Prefix conjugation hp’l (C-stem)
*35 Prefix conjugation ’p’l (C-stem)
*36 Prefix conjugation šp’l (C-stem)
*37 Prefix conjugation htp’l (Gt-stem or Dt-stem)
*38 Prefix conjugation ’tp’l (Gt-stem or Dt-stem)
*39 Specifically Arabic verbal forms (tafa’ala, infa’ala, etc.) prefix conjugation

[Other grammatical categories start with 4 or 5]
*40 Other verbal form unspecified
*41 Infinitive
*42 Imperative
*43 Active participle sg.
*44 Active participle pl.
*45 Passive participle sg.
*46 Passive participle pl.
*47 Arabic form of active participle
*48 Arabic form of passive participle
*49 AVAILABLE TAG
*50 Adjective unspecified
*51 Adjective sg.
*52 Adjective pl.
*53 Gentilic adjective sg.
*54 Gentilic adjective pl.
*55–*59 AVAILABLE TAGS

[All pronouns start with 6]
*60 Pronoun unspecified
*61 Independent personal pronoun
*62 Enclitic personal pronoun sg.
*63 Enclitic personal pronoun pl.
*64 Determinative relative pronoun
*65 Indefinite relative pronoun
*66 Demonstrative pronoun masc. sg.
*67 Demonstrative pronoun fem. sg.
*68 Demonstrative pronoun pl. (masc. & fem.)
*69 AVAILABLE TAG

[All particle and adverbs start with 7]
*70 Particle, adverb, unspecified
*71 Preposition
*72 Conjunction
*73 Interjection
*74 Negative particle
*75 Genitive particle [e.g. npš dy so-and-so]
*76  Arabic definite article
*77  Adverb
*78  Other particle
*79  AVAILABLE TAG

**[All proper nouns start with 8]**
*80  Name unspecified
*81  Man’s personal name
*82  Woman’s personal name
*83  Undifferentiated personal name
*84  King’s name
*85  Queen’s name
*86  Name of royal child
*87  Divine name
*88  Non gentilic group name
*89  Place name

**[Other word categories start with 9]**
*90  Time, numbers and miscellaneous unspecified
*91  Month name
*92  Name of era
*93  Number in numerals
*94  Number in letters
*95  Foreign substantive transcribed
*96–*97 AVAILABLE TAGS
*98  To be tagged
*99  Indeterminate