

Lexikografická databáze beduínských dialektů

Informace pro uživatele

Jakub Koláček, Adéla Provazníková, Petr Felčer

Kontakt: jakub.kolacek@seznam.cz

30. 12. 2016

Tento výstup vznikl v rámci projektu "Paměť jazyka, svědectví literatury: analýzy literárního a lingvistického prostoru napříč žánry a kontexty" podprojektu "Tvorba lingvistické lexikální databáze z terénních záznamů lexika arabských Šammarů z pozůstalosti Aloise Musila" na rok 2016 řešeného na Filozofické fakultě Univerzity Karlovy v Praze z prostředků Specifického vysokoškolského výzkumu.

Úvod

Tato databáze byla vytvořena ze souboru 5415 karet obsahujících lingvistický materiál nalezených v pozůstalosti Aloise Musila.¹ Tento soubor je jedním z prvních dochovaných dokladů o jazyce beduíňů srovnatelného rozsahu a lze jej využít pro studium beduínských dialektů ze synchronního i diachronního hlediska. Soubor se převážně týká dialektu beduínského kmene Rwalů. Část obsažených lexémů byla zahrnuta do Musilovy monografie *The Customs and Manners of Rwala Bedouins* vydané v r. 1928 v New Yorku.²

Cílem přepisu bylo, aby databáze umožnila vyhledávání a třídění obsahu podle různých kritérií (tedy využití předností elektronizace) při souvztažném zachování původní podoby Musilových záznamů, která je často důležitá pro jejich interpretaci.

Databáze je volně k dispozici pro další editaci a rozšiřování. Autoři budou vděční za veškeré ohlasy, připomínky či kritiku. Spolu s případnými dotazy je adresujte na výše uvedenou kontaktní adresu.

Popis databáze

1. Obecné informace

- Do databáze byl přepsán každý jednotlivý lístek Musilovy kartotéky. Pakliže lístek zahrnuje více informací, je dále rozdělen do vícera jednotek.
- Jednotky databáze (o jejich specifikaci viz níže) jsou uvedeny v řádcích, sloupce obsahují logické kategorie pro třídění a další práci se záznamy.

- Základní jednotkou databáze je konstruovaná sémantická jednotka – může se jednat o lexém, frazém či syntagmém v určitém gramatickém tvaru, tak jak je zaznamenán v původním souboru dat. Tato jednotka je z původního datového záznamu (tedy katalogového lístku) extrahována. Zásady, kterými jsme se řídili při konstrukci těchto jednotek, budou uvedeny dále. Další jednotku databáze představuje původní katalogový lístek ve formě čísla souboru digitalizovaného záznamu. Každý lístek může jako jednotka databáze obsahovat jednu či více sémantických jednotek. Databáze je strukturována tak, že zachovává informace o příslušnosti všech záznamů k originálním katalogovým lístkům. Původní struktura Musilových záznamů se tak dá z databáze plně rekonstruovat. K třídění může sloužit taktéž kořen a Musilův alfanumerický kód uvedený v hlavičce lístku.
- Kategorie se dělí do tří skupin:
 - **Hlavička** lístku nesoucí informaci o kořeni, alfanumerický kód a číslo lístku. Tyto kategorie jsou označeny zelenou barvou.
 - **Sémantická jednotka a související data** (oranžová).
 - **Kompletní přepis tematického obsahu lístku** (vyjma hlavičky a arabského záznamu) zahrnujícího typicky (1) transkripci arabského výrazu/fráze a (2) vysvětlení jejího významu, popřípadě i další informace a další lexikum (které je rovněž převedeno do sémantických jednotek).
- Lístky jsou roztrženy podle jazyka vysvětlení na českojazyčné a německojazyčné. Překlad z němčiny do češtiny není v databázi zahrnut.
- U části položek jsou některé údaje (např. kořen slova) v databázi doplněny, ačkoliv nebyly součástí původního záznamu. Tato doplnění slouží pro lepší třídění položek a jednodušší práci s výslednou databází. Zároveň je zachována informace, že se jedná o nepůvodní, doplněné údaje. Podrobnosti jsou uvedeny níže.
- V případě transkribovaných arabských lexémů byly k přepisu Musilova rukopisu a strojopisu použity znaky, které se nejvíce blíží originální podobě na lístcích:
 - konsonanty (dle pořadí v arabské abecedě): ’, b, t, ṭ, ġ, ḥ, ḥ, d, ḍ, r, z, s, š, ṣ, ḏ, ṯ, z, ‘, ṛ, f, k, k, l, m, n, h, w, j, ’;
 - vokály: a, â, ä, e, ê, i, î, u, û, o, ô.

- Znaký pravděpodobně odpovídají arabským grafémům podle Tabulky 1, uvedené v knize *The Manners and Customs of Rwala Bedouins*.

Tabulka 1

TRANSLITERATION

Line (-) or bow (˘) under the letter indicates an aspirated, a dot (.) an emphatic, and the sign (˙) over the letter a palatal pronunciation.

ʾ = ʾ	z = ز	k = ك
b = ب	s = س	ẓ = ڙ (like z in azure)
t = ت	š = ش (sh)	k = ك (like k in like)
ṭ = ٲ (ts)	ṣ = ص (sz)	č = چ (ch)
ḡ = ڭ (like g in gem, or dsh, dj, or y in yoke)	ḍ = ڌ (dh, dah)	l = ل
h = ح	ṭ = ٲ	m = م
ḥ = ح	ẓ = ڙ (zh)	n = ن
d = د	˙ = ع (strong, guttural sound)	h = ه
ḍ = ڌ (dz)	ṛ = ر (ghr)	w = و
r = ر	f = ف	j = ڃ (y in yoke)

2. Specifikace jednotlivých kategorií

A. Hlavička

I. LÍSTEK

- Jedná se o prosté číslo souboru bez přípony, pod kterým je uložen scan původního archivovaného lístku.
- Číslo lístku je uvedeno vždy u všech zahrnutých sémantických jednotek (je-li jich na lístku více, následuje po sobě více řádků se stejným číslem).

II. KOŘEN ARABSKY

- Jedná se o arabský zápis kořene uvedený v některých případech Musilem v hlavičce lístku.
- Kořen je v databázi uveden vždy primárně v podobě uvedené Musilem, buď přímo, nebo na základě transkribovaného zápisu. Není-li kořen na lístku původně uveden (typicky u dalších sém. jednotek na lístku), je doplněn autory. Jednotlivé případy jsou odlišeny (viz níže).

- Pokud se jedná o kořen se shodným druhým a třetím radikálem, opakující se konsonant je uveden pouze jednou a bez *šaddy* (znaménka pro zdvojení).
- Poněkud problematické jsou případy kořenů s druhým a třetím slabým vokálem, které Musil zapisuje nesystematicky – např. حشو i حشا، حشى . Databáze se drží původních záznamů a zápis není jednocen, proto je při vyhledávání v těchto případech třeba zahrnout všechny varianty které připadají v úvahu.
- Pokud kořen nelze uvést, je uveden znak „X“. Případné dodatečné informace jsou uvedeny v poznámce.

III. CHARAKTERISTIKA ARABSKÉHO KOŘENE

- Jde o kód upřesňující původ záznamu v kategorii II.
- 1 = kořen je uveden Musilem v této podobě
- 2 = kořen je odvozen z Musilovy transkripce
- 3 = kořen nebyl uveden a je doplněn badatelem
- CS = kořen nelze uvést, neboť se jedná o cizí slovo

IV. FONETICKÝ PŘEPIS KOŘENE

- Fonetický přepis kořene je na lístcích uveden charakteristicky tam, kde arabský záznam chybí.
- Je-li na lístku uveden, je doslovně zaznamenán, v opačném případě nikoliv.
- Musil využívá speciální znak pro označení prodloužení konsonantu u sloves se třetím radikálem slabým. Tento znak zachováme jako „,““ (viz informace o transkripci).

XII. ALFANUMERICKÝ KÓD

- Jedná se o kód uvedený Musilem v hlavičce většiny lístků.³
- Je zaznamenán v doslovné podobě s výjimkou některých korekcí pro sjednocení záznamu:
 - Mezi písmennou (patrně značí soubor) a číselnou (patrně značí položky v něm) částí kódu je vždy mezera (Wetzstein, Sprachliches 12, 23 ; A.D. 352 ; Zv. 18, 21 ; F 33 ; H 11).
 - Sestává-li číselná část ze dvou čísel oddělených čárkou, následuje za čárkou mezera.

- U indexovaných odkazů je uvedeno prosté písmeno a číslo oddělené mezerou.
- Je-li záznam zakončen tečkou, není uvedena.

B. Sémantická jednotka a související informace

V. ARABSKÁ SÉMANTICKÁ JEDNOTKA

- Jde o základní jednotku databáze. Rozdělením souboru dat na sémantické jednotky je umožněna efektivní práce s daty – jejich rychlé prohlížení, třídění a prohledávání.
- Jedná se o **konstruovanou jednotku**, vyňatou ze zbytku údajů na lístku. Výběr jednotek a jejich definice představuje dodatečné třídění nad rámec původního datového souboru.
- Jako sémantická jednotka byl do databáze zaznamenán každý arabský lexém či syntagmém identifikovaný na lístku. Postup se řídil zásadou požadující, aby každý plnovýznamový⁴ lexém, který je na lístku přítomen, byl zahrnut do samostatné jednotky a byl identifikován se svým slovním kořenem. Pokud je předmětem původního Musilova záznamu frazém či syntagmém složený z více plnovýznamových slov, je nejprve zachován a poté rozdělen, a to až na úroveň která odpovídá jednomu slovnímu kořeni.
- Původní podoba transkripce je zachována, pouze velké znaky jsou převedeny na malé.

VI. CHARAKTERISTIKA SÉMANTICKÉ JEDNOTKY

- Tato kategorie specifikuje kvalitu sémantické jednotky a lze ji využít k dalšímu třídění.
- PF = předložková fráze
- NF = nominální fráze
- VF = verbální fráze
- S = samostatné slovo

VII ARABSKÝ ZÁZNAM

- Zahrnuje arabský záznam uvedený Musilem na původním lístku, který je ze zbytku informací vyňat.
- Nezahrnuje kořen, který je většinou jasně odlišen.
- Tyto případy jsou relativně ojedinělé.

VIII MNOŽNÉ ČÍSLO

- Je-li Musilem explicitně uvedeno množné číslo, není zahrnuto jako samostatná sémantická jednotka a je zařazeno do této kategorie. Přepis se řídí stejnými zásadami, jako v případě sémantické jednotky.

C. Přepis tematické části lístku

IX DOSLOVNÝ OBSAH LÍSTKU V ČESKÉM JAZYCE BEZ HLAVIČKY

- Do kategorie byla zaznamenána celá tematická část lístku zahrnující arabský lexém či syntagmém a vysvětlení významu v českém jazyce, a to v doslovné podobě, vyjma ojedinělých oprav zjevných překlepů a interpunkce. Zahrnuté jsou i přeškrtnuté části záznamu a ruční vpisky do strojopisu (pakliže je lze považovat za relevantní).
- V případě že Musil rozdělil informace do několika řádků, bylo k jejich oddělení použito zpětné lomítko (\). K označení chybějícího či nečitelného záznamu byly využity tři tečky (...) nebo otazník (?).
- Kategorie se systematicky kryje s kategorií I, u lístků s totožným číslem je uveden vždy totožný doslovný obsah.
- Záznamy v této kategorii lze využít k fulltextovému vyhledávání na základě českého vysvětlení významu.

X CHARAKTERISTIKA OBSAHU LÍSTKU

- Slouží k odlišení lístků vysvětlujících význam v češtině, nebo v němčině.
- Česky = Č
- Německy = N
- Záznam je nečitelný / neobsahuje relevantní informaci = X

XI DOSLOVNÝ OBSAH LÍSTKU V NĚMECKÉM JAZYCE BEZ HLAVIČKY

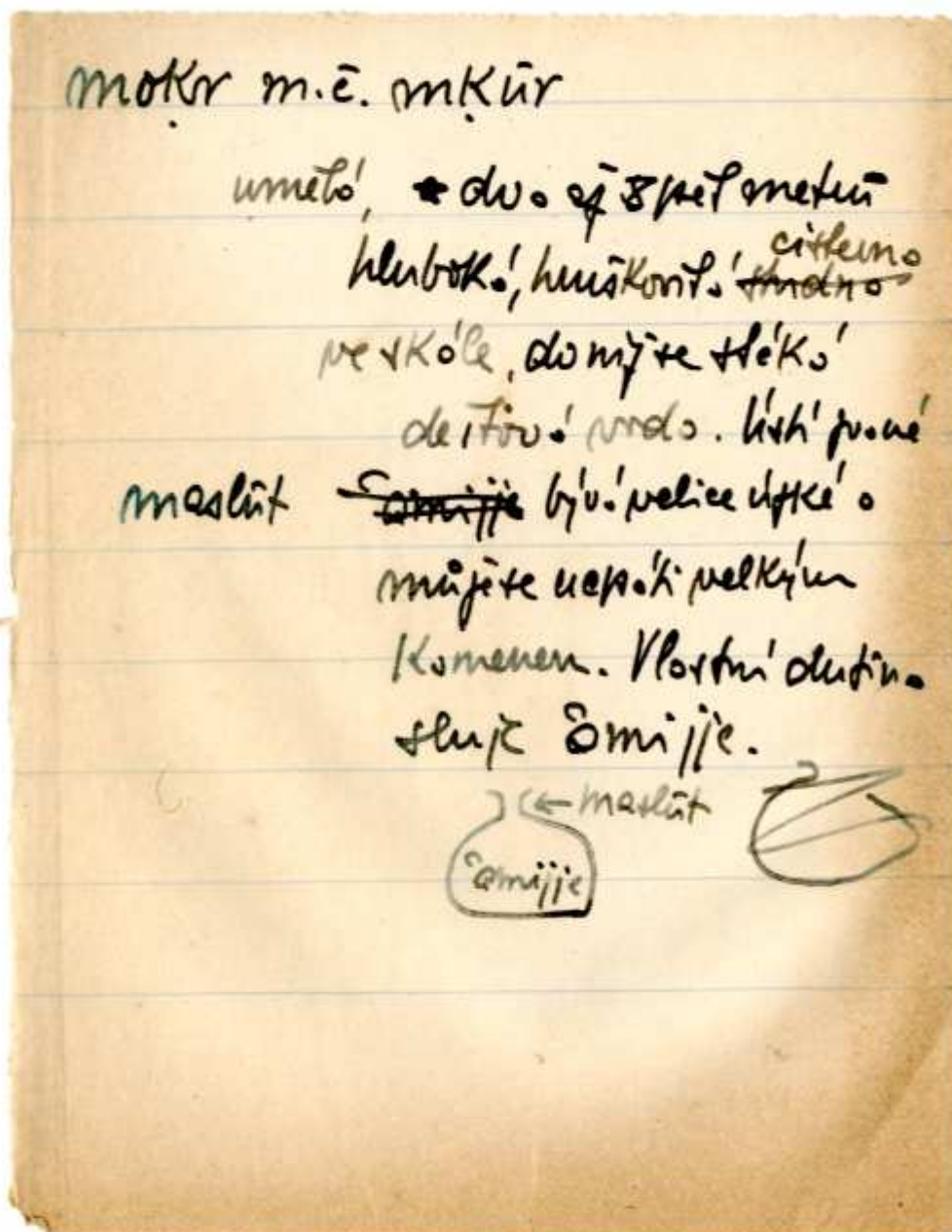
- Jedná se o obdobu kategorie IX pro německojazyčné lístky, které byly v databázi zaznamenány do zvláštní kategorie.

Závěr

Do zmíněných kategorií jsou zaznamenány všechny relevantní údaje ze všech lístků původního souboru. Při přepisu bylo identifikováno celkem 8634 sémantických jednotek, z nichž 6550 je formálně odlišných, a asi 2500 různých slovních kořenů. Přibližně 560 lístků obsahuje vysvětlení v německém jazyce (výjimečně je záznam smíšený) a zhruba 100 lístků bylo označeno jako defektních (význam zde chybí, nebo jsou obsaženy irelevantní

informace). Dodatečné informace k jednotlivým jednotkám databáze jsou uvedeny ve sloupci *Poznámky*.

Obrázky 1 (č. 0056), 2 (č. 3231) a 3 (č. 3359) ilustrují podobu původních Musilových záznamů.

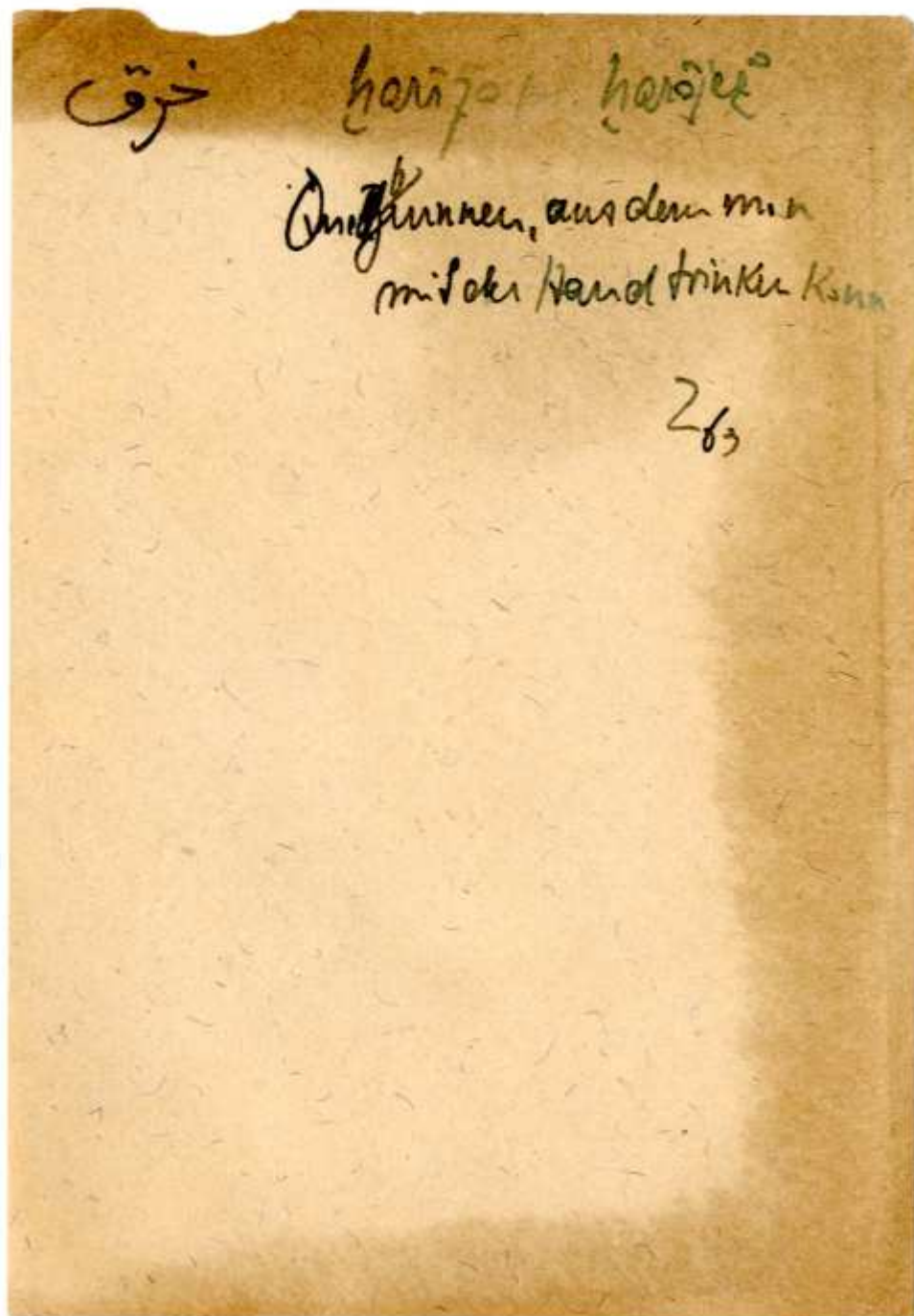


Obr. 1

Hâr.

Zv 5,10

do fimeñca-
Hwâr - velbloudě ~~v prvé~~ roce.



Obr. 3

1Pro podrobný popis nálezů viz Sadílek, Tomáš, Pavel Žďárský a Petr Přebinda: Arabsko-český slovník beduínských výrazů v pozůstalosti Aloise Musila. *Časopis Národního muzea, řada historická*, 182/1–2 (2013), s. 19–28.

2Viz Musil, Alois. *The Manners and Customs of the Rwala Bedouins*. New York, American Geographical Society of New York, 1928.

3Kód ve tvaru *Zv.* + *číselný údaj* odkazuje na jednotlivé kapitoly Musilovy práce *The Manners and Customs of Rwala Bedouins* (Musil, Alois. *The Manners and Customs of the Rwala Bedouins*. New York, American Geographical Society of New York, 1928), pro kterou tyto lístky patrně sloužily jako podklad. Kód *Wetzstein, Sprachliches* + *číselný údaj* odkazuje na jednotlivé strany práce *Sprachliches aus den Zeltlagern der syrischen Wüste* pruského konzula v Damašku J. G. Wetzsteina. (Wetzstein, Johann Gottfried. *Sprachliches aus den Zeltlagern der syrischen Wüste*. Leipzig, 1868.) V případě dalších kódů se lze domnívat, že je situace podobná – jedná se buď o vnitřní řazení, případně odkazy na jiné Musilovy práce (?A.D.=Arabia Deserta, Neğd=Northern Neğd, H=The Northern Heğâz?), nebo odkaz na určitou jazykovědnou práci, kterou Musil komentuje (články J.J.Hesse v časopise *Der Islam*).

4Tedy nikoliv předložka, částice, zájmenný sufix, člen.