

Frequency of Using Najdi Arabic Words Among Saudi College Male Students

Sayed M. Ismail¹, Nasser Rashid Alshayhan², Salwa Alwafai³ & Bacem A. Essam⁴

¹ Assistant Professor of Translation Studies, Prince Sattam Bin Abdulaziz University, Saudi Arabia

² Assistant Professor of Arabic Language and Literature, Prince Sattam Bin Abdulaziz University, Saudi Arabia

³ Assistant Professor of Linguistics and Translation Studies, Department of English-College of Arts, Imam Abdulrahman Bin Faisal University, Al-Baath University (previously), Saudi Arabia

⁴ Ain Shams University, Egypt

Correspondence: Sayed M. Ismail, Prince Sattam Bin Abdulaziz University, Alkharj Community College. Abu Dajan Street-Al-montzha District. P. O. Box 1866, Alkharj 11942, Riyadh, Saudi Arabia. E-mail: a.ismail@psau.edu.sa

Received: December 11, 2018 Accepted: January 15, 2019 Online Published: January 29, 2019

doi:10.5539/ijel.v9n2p24

URL: <https://doi.org/10.5539/ijel.v9n2p24>

Abstract

The study of dialects may be subsumed under the very broad rubric of colloquialism which comes at the bottom of the formality versus informality scale. We focus on the Najd dialect perception, as the central dialect in Saudi Arabia, among Saudi male college students. By conducting two experiments, questionnaires and follow-up semi-structured interviews, on 137 male students, user-based frequencies of the top 50 Najdi words are generated. The second phase aims at semantically categorizing the top 50 content words so that conclusions can be drawn about the inclination of using Najdi words among the college students. Results show that the categorization of the retrieved 50 Najdi words, according to the part of speech, demonstrates that the most applauded Najdi Arabic words are verbs and adjectives. Synonyms are even retrievable from this method of compilation. Nouns are the most resistant part of speech at the morphological level.

Keywords: Najdi Arabic, colloquialism, cultural cognition, language resources, lexicography, sociolinguistics

1. Introduction

1.1 Problem Statement

The poor resourcing of Arabic language and its dialects, unavailability of competent morphological analyzers and syntactic parsers, specificity of the informal discourse and user-generated content, writing inconsistency of users and the dynamic sociolinguistic shifts are obstacles that impede any synchronic understanding of language dynamics and change. This seems to be true especially for the Najdi dialect; compared to the resources and tools applicable on the standard language variety and to the plethora of the enhanced tools for the Indo-European languages. This study, therefore, addresses such a challenge by introducing a dialectal frequency-based lexicon of Najdi words and their English translation. To validate our proposed lexicon, we investigate the familiarity of 137 Saudi college male students with the collected set of words. This effort aims to answer two questions. First, what are the most frequent Najdi content words according to the Saudi male college students? Second, how accurate is the automatic detection of dialectal vocabulary and of their frequency?

1.2 Review of Literature

The psycholinguistic profiling of Najdi dialect is scarcely addressed. Ismail (2017) studied the translation challenges of rendering Najdi proverbs into English. He concluded that the process of translating Najd dialect necessitates understanding the external real world in which a dialectal word lives and articulates because dialect words, and their senses, derive meanings from their cultural realities and from their historical context. Alatawi (2015) analyzed the structural, socio-pragmatic, and psycholinguistic analysis of code switching in Arabic TV programs either monolingually among the major dialects (Najdi, Hejazi, Janubi, Shamali & Shargiah) or bilingually between Arabic and English. Ingham (1994) introduced a bilingual Najdi Arabic-English lexicon. However, his study lacked rigorous criteria of inclusion. Therefore, the specificity of the generated wordlist, and

lemmas, was not acceptable. This highlights the significance of studying the Najdi dialect as objectively as possible to define the most frequent words and what they may tell about the original culture. In a similar vessel, lexicographic effort has been recently directed toward compiling frequency-based dictionaries for core Arabic vocabulary (Buckwalter & Parkinson, 2014) and even for word sketches, collocates and thematic lists of American English (Davies & Gardner, 2013).

Recently, StimulStat, a frequency-based psycholinguistic resource for Russian, is proposed to include approximately 1.7 million word-forms, lemmatized into 52,000 lemmas. Frequency, length, uniqueness point, orthographic, phonological representations (stress, syllabic structure, phonemic transcription), grammatical features, semantic relations (homonymy, polysemy, synonymy), typical age of acquisition and imageability are all covered. StimulStat was built on three printed dictionaries, of which was a frequency lexicon, and a national corpus (Alexeeva et al., 2018).

Drawing on the organization of the mental lexicon, psychologists and lexicographer studied psychologically realistic models of polysemy and synonymy in compilation of general formal dictionaries (Jorgensen, 1990; Miller et al, 1998; Wilks, 2008) and colloquial lexica (Arenas, 2018; Collins & Yao, 2018; Essam, 2017; McCrae, 2017).

2. Theoretical Background

Other than classical Arabic and modern standard Arabic, there are several dialect groups. The major dialectal groups are the Egyptian, Arabian Peninsula, Maghrebi, Sudanese, Mesopotamian, Levantine, Andalusian dialectal groups. Dialect expressions are informal, temporal, and oral. They are also dynamic, novel, often colorful, and humorous and aim either to establish a social identity for the speaker or to make strong impression upon the hearer (Mattiello, 2009). It is thus intelligible and recognizable for the people who share the geographical borders to speak the same dialect as a means for reflecting the identity of the speakers. Considered ambiguous, figurative and specific, the connotation of dialectal expressions is comprehensible for the people who use such expressions in their day-to-day life. Thus, exclusivity is a major feature of dialectal expressions. Playfulness is a common feature that designates the dialectal expressions. In many cases, the dialectal expressions are figurative, as they use metonymy and pun in communicating their intended messages. Dialectal expressions are creative and innovative. They are fallen under the process of permanent reproduction and reconstruction. Spatial and socioeconomic realities stain Arabic dialectal expressions with totally different connotations from the modern standard Arabic (Holes, 2018). The culture-bound elements refer to items, tools, instruments, proper names, food items related to specific culture and specific time. The problem of translating culture bound elements is threshold (Mattiello, 2009).

Najdi dialect is an oral form of speech, which is transferred across generations through imitation and repetition. It is transferred through folklore and oral forms of poetry, like Nabati poetry. In addition, the names of traditional industries, artifacts, crafts, and cultural heritage are realized in dialects. The formation of dialectal words and expression is time bound as well as cultural bound. That is to say, dialect words rapidly change over time because they derive their meanings and understanding from their surrounding world (Holes, 2018). Since Najd region hosted the capital city of Saudi Arabia, Riyadh, there has been a clear shift in the socio-economic realities of Saudi community. The nomadic communities had been replaced by stable communities living in well-organized and modern cities. That is to say, urbanization prevailed over Najd region. The Saudi government began establishing very modern system of education as modern schools and universities have been built in Saudi Arabia. Multiple dialect words are borrowed from Pakistani, Indian, Bengali and English. Today, Najd region is no longer such an isolated desert land. On the contrary, it has become an international city that attracts people from all over the world. This whole issue has clearly been reflected in being the central dialect in Saudi Arabia.

Some dialect expressions are not clear and incomprehensible, which requires interpreting and understanding them in MSA. Given that the intra-lingual concept of translation is no longer valid for translating dialects, because translators fail to retrieve synonyms for the dialectal terms either in the classical Arabic language or in MSA, lexicographic effort is ushered towards computationally generating and compiling frequency-based specific language resources.

3. Method

3.1 Participant Characteristics

The participants were voluntarily recruited from Prince Sattam Bin Abdulaziz University at their sophomore year. They were all male native speakers of Arabic who studied language and translation. The participants were asked to report any tribal affiliation if indicated. The mean age of the students was 19 years (range: 18 to 20). A

link to the online survey, using SurveyMonkey, was shared. Those who did not wish to fill in the electronic form were given printed versions of the questionnaire. After responding to the questionnaire, the volunteering students were interviewed. In the interviews, questions were structured to measure the familiarity of the students with the context of the Najdi Arabic words in which they normally live. Put simply, participants were asked to suggest any sentence they usually hear, which contains the investigated Najdi words. A total of 137 responses were collected.

3.2 Study Material

We extracted wordlists from two corpora. The first one, TenTen15, was a ready-made searchable corpus, retrieved from the web, and did include both modern standard Arabic and several dialects. The second corpus was generated from online tweets published by Saudi citizens using API bootstrapping. Wordlists were computationally retrieved and sorted on frequency. A list of the 50 most frequent Najdi words were exported into our questionnaire and were sorted on the ranking of the corpus-driven findings. Both electronic and printed forms of the questionnaire were produced to match the preferences of the respondents.

3.3 Study Procedures

This study conducted three experiments. First, we computationally compiled most frequent Najdi content words from general and specific contemporary corpora. The frequency of occurrence of each lemma in the corpus and concordance were also retrieved. Second, a questionnaire, which was composed of a 5-Likert-scale question about the frequency of every Najdi word, was designed and tested on 137 male students. All students were sophomore Saudi college students. Third, interviews were conducted to investigate the familiarity of the respondents with the studied Najdi words. Statistical *t*-test was calculated to compare the human-based evaluation to the automatic computational compilation.

4. Results

In this section, we describe the results of the three experiments we have conducted. For the first experiment, the list of retrieved most frequent Najdi content words and their frequencies are provided. The functional words and non-words were excluded. Table 1 displays the frequency of the usage of the 50 topper Najdi Arabic words (NAW) among the studied population. Translation was rendered according to the concordance of each NAW. The Likert scale, used for respondents to evaluate the frequency of using each NAW, was sorted from the most frequent (daily heard NAW) to the least frequent (rarely heard). Respondents demonstrated familiarity with all the studied NAWs.

Table 1. Frequency of the usage of the 50 topper Najdi Arabic words (NAW) among the studied population

NAW	NF	Translation	Daily	MTW	MTM	OM	Rarely
هَج	0.0335	Escape	28.5	28.5	21.2	5.7	16.1
هَك	0.0089	Lie	24.8	25.7	28.5	7.3	13.7
تَقْطِن	0.0096	Investigate	22.6	22.6	21.2	15.3	18.3
نَظَل	0.0065	Throw	19	19.6	21.2	11.7	28.5
مَنْس	0.0063	concealing	19	19.7	19	9.5	32.8
عَنْز	0.0059	Lean	25.6	24.8	15.3	16.8	17.5
هَنْيَه	0.0053	Here	17.6	16.8	22.6	15.3	27.7
زَبَن	0.0047	Lodging	28.5	28.5	21.2	5.7	16.1
نِيْم	0.0047	Sleepy	24.8	24.8	22.6	6.6	21.2
سَسْتَم	0.0034	System	28.5	28.5	27	7.3	8.7
مَارَج	0.0024	Rotten	33.6	33.6	16.8	8	8
عَطْن	0.0020	Gimme*	24.8	24.8	22.6	8.8	19
خَرَاط	0.0017	Liar	21.9	21.9	19.7	8.8	27.7
زَنْط	0.0017	Suffocate	27	27	19	9.5	17.5
اِبْخَص	0.0015	Expert	21.9	21.9	19	10.9	26.3
اَنْهَج	0.0015	Faster	20.4	20.4	18.2	10.9	30.1
غَضَارَة	0.0014	Bottle	21.9	21.9	19	11.7	25.5
زَمِين	0.0012	Likely	23.4	23.4	21.2	13.1	18.9
بِرْطَن	0.0011	Talk much	24.8	24.8	22.6	13.1	14.7
شُرَوَاك	0.0009	Like you	26.3	26.3	24.1	13.9	9.4
يَهَج	0.0008	Escape	27.7	27.7	22.6	14.6	7.4
لَقْف	0.0006	Interfere	24.1	24.1	21.9	15.3	14.6
المِيَانَة	0.0006	Sanction	28.5	28.5	20.4	13.1	9.5

هبيلة	0.0005	Insane	29.2	29.2	19	10.9	11.7
هبره	0.0005	Mountain	27.7	27.7	17.5	8.8	18.3
برطم	0.0004	Lips	24.1	24.1	15.3	10.2	26.3
معصي	0.0004	Refuse	22.6	22.6	18.2	6.6	30
متروس	0.0004	Stuffed	24.8	24.8	19	8	23.4
مملوحة	0.0002	Beautiful	49	19	17.5	11.3	3.2
شقردي	0.0002	brave	24.8	24.8	16.8	7.3	26.3
نفج	0.0002	Escape	33.4	23.4	13.1	8	22.1
لحجي	0.0002	Peasant	27.7	27.7	10.2	8.8	25.6
مصرف	0.0002	Reckless	24.8	24.8	9.5	13.1	27.8
مطنوخ	0.0002	Rich	28.5	28.5	8.8	9.5	24.7
مشتت	0.0001	Angry	39	19	10.9	13.1	18
دبش	0.0001	Cow	26.3	26.3	13.9	10.9	22.6
كمخة	0.0001	Dummy	21.2	21.2	16.8	10.9	29.9
سبيهه	0.0001	Insane	20.4	20.4	19.7	5.8	33.7
دوشق	0.0001	Mattress	21.2	21.2	22.6	5.8	29.2
سرويس	0.0001	Plate	49	19	13.9	5.8	12.3
هجوالة	0.0001	Gaudy	41.2	21.2	13.1	5.8	18.7
يهوجس	0.0001	Think	42.6	22.6	12.4	5.8	16.6
قروشة	0.0001	Troubling	34.1	24.1	10.9	5.8	25.1
ملجلج	0.0001	Upsetting	25.5	25.5	10.9	5.8	32.3
يحتري	0.0001	Wait	27	27	10.2	5.8	30
هقوة	0.00001	Believe	28.5	28.5	11.7	5.8	25.5
ازبن	0.00001	Conceal	30.7	30.7	13.9	5.8	18.9
يسعبل	0.00001	Droll	31.4	31.4	15.3	5.8	16.1
يزغت	0.00001	Gormandize	25.5	25.5	19.7	10.9	18.4
يكمخ	0.00001	Hit	28.5	28.5	21.2	5.8	16
بريعصي	0.00001	Lizard	24.8	24.8	28.5	7.3	14.6
مطرسة	0.00001	mix	22.6	22.6	21.2	15.3	18.3
مطارس	0.00001	Mixture	29	19	21.2	11.7	19.1
مفصخ	0.00001	Naked	29	18	19	10.5	23.5
قرمبع	0.00001	Relic	24.8	24.8	15.3	16.8	18.3
بويق	0.00001	Disregard	16.8	16.8	22.6	15.3	28.5
طوفرية	0.00001	Plate- tray	28.5	28.5	21.2	5.8	16
مزلب	0.00001	Reckless	24.8	24.8	22.6	6.6	21.2
ذعرتني	0.00001	Scared	28.5	28.5	27	7.3	8.7
منجطل	0.00001	Side-sleeping	33.6	33.6	16.8	8	8
يتن	0.00001	Smoke	24.8	24.8	22.6	8.8	19
مردغة	0.00001	Stroke	21.9	21.9	19.7	8.8	27.7
يزوزي	0.00001	Whisper	27	27	19	9.5	17.5

Note. Abbreviations: MTW: Multiple times weekly; MTM: Multiple times monthly; NF: Normalized Frequency; OM: Once monthly. * Short form for "give me".

The categorization of the studied 50 Najdi words, according to the part of speech, demonstrates that the most applauded NAWs are verbs and adjectives. Synonyms are even retrievable from this method of compilation (e.g., هبيلة and سبيهه). Nouns are the least changing form morphologically. However, unpaired *t*-test and ANOVA test revealed a statistical significance between the frequencies retrieved computationally and the frequencies scored by the students. Post-hoc Tukey test showed that verbs and adjectives were of greater familiarity to respondents than of what numbers predicted.

Table 2. Frequency of the usage of the 50 topper Najdi Arabic words (NAW) among the studied population

Verbs		Adjectives		Nouns		Adverbs	
هَج	Escape	مُتَدَس	concealing	زِين	Lodging	هَنِيه	Here
هَك	Lie	نِيم	Sleepy	سِسْتَم	System	زَمِين	Likely
تَقْطَن	Investigate	مَارَج	Rotten	اِبْخَص	Expert	مَنْجَطَل	Side-sleeping
نَطْل	Throw	خَرَاط	Liar	غَضَارَة	Bottle		
عَنْز	Lean	اِنْهَج	Faster	اَلْمِيَانَة	Sanction		
عَطْن	Gimme	يِرْطَن	Talkative	هَبْرَه	Mountain		
زَنْط	Suffocate	شُرَوَاك	Like you	بِرْطَم	Lips		
يَهَج	Escape	هَبِيلَة	Insane	لَحْجِي	Peasant		
لَقْف	Interfere	مُتْرُوس	Stuffed	دَبْش	Cow		
مَعْصِي	Refuse	مَمْلُوحَة	Beautiful	دُوشَق	Mattress		
نَفْج	Escape	شُقْرَدِي	brave	سُرُوس	Plate		
يَهْوَجْس	Think	مَصْرَقَع	Reckless	بِرْيَعْصِي	Lizard		
يَحْتَرِي	Wait	مَطْنُوح	Rich	مَطَارِيس	Mixture		
هَقْوَة	Believe	مُشْتَب	Angry	قَرْمِيع	Relic		
اَزِين	Conceal	كَمْخَة	Dummy	طُوفْرِيَة	Plate/tray		
يَسْعَبَل	Droll	سَبِهَة	Insane				
يَزْغَت	Gormandize	هَجُولَة	Gaudy				
يَكْمَح	Hit	قُرُوشَة	Troubling				
مَطْرَسَة	mix	مَلْجَلَج	Upsetting				
يُوق	Disregard	مَفْصَح	Naked				
يَتْن	Smoke	مَزْلَب	Reckless				
مَرْدَغَة	Stroke	ذَعْرَتْنِي	Scared				
يَزُورِي	Whisper						

5. Discussion

Dialect expressions are often informal, temporal, and dynamic. They are used for establishing a social identity or for making strong impression upon the hearer (Mattiello, 2009). Najdi dialect is an oral form of speech, which is transferred across generations through imitation and repetition. It is transferred through folklore and oral forms of poetry, like Nabati poetry. In addition, the names of traditional industries, artifacts, crafts, and cultural heritage are realized in dialects. Notwithstanding, some dialect expressions are not clear and incomprehensible, which requires interpreting and understanding them in MSA. Psycholinguistic lexicographic effort is ushered towards computationally generating and compiling frequency-based specific lexicon for addressing the most frequent Najdi dialect expressions.

In our proposed lexicon, we excluded the functional words because they are systematically used in all languages. Therefore, the incidence of its use is not of paramount importance, unlike the content words. The majority of content words denote thematically motion verbs, size description, emotion regulation and culture-specific designations. The results retrieved from Hanson et al. (2001) go semantically hand in hand with our results. Such result reported a greater predilection of frequently using words which express activities, persons (patients and agents) and the products of human activities.

To the best of our knowledge, this is the first study which aims at compiling a frequency-based Najdi Arabic-English lexicon. For the accuracy of NAW detection, the computational retrieval was successful at retrieving the topper NAWs. However, this technique was more accurate in retrieving the frequency of nouns and adverbs than in measuring the frequency of verbs and adjectives. However, computational bootstrapping seems more effective in generating a customized search because the top-down approach could not retrieve as much words as required. We therefore recommend integrating both techniques for compiling dialectal lexica.

6. Conclusion

Frequency-based dialectal lexical are useful for building up language resources. The specificity and sensitivity of the retrieved lexica depends greatly on the adopted method of compilation. We recommend the interoperability between automatic detection of the seeding words and the psycholinguistic experimentation and validation of the generated vocabulary. We also recommend the social media streams as a rich medium for generating colloquial dictionaries. Contribution of similar studies to machine translation and to building language resources would be made. This lexicon can be integrated among other dialectal varieties of Arabic to enrich such a poorly resourced language.

References

- Alatawi, E. (2015). *Structural, socio-pragmatic, and psycholinguistic analysis of codeswitching in Arabic TV programs*. Doctoral dissertation, Northeastern Illinois University.
- Alexeeva, S., Slioussar, N., & Chernova, D. (2018). StimulStat: A lexical database for Russian. *Behavior Research Methods*, 50(6), 2305–2315. <http://doi.org/10.3758/s13428-017-0994-3>
- Arenas, E. S. (2018). Exploring Pornography in Widad Benmoussa's Poetry Using LIWC and Corpus Tools. *Sexuality & Culture*, 22(4), 1094–1111. <http://doi.org/10.1007/s12119-018-9513-7>
- Buckwalter, T., & Parkinson, D. (2014). *A frequency dictionary of Arabic: Core vocabulary for learners*. Routledge. <https://doi.org/10.4324/9780203883280>
- Collins, P., & Yao, X. (2018). Colloquialisation and the evolution of Australian English. *English World-Wide*, 39(3), 253–277. <https://doi.org/10.1075/eww.00014.col>
- Davies, M., & Gardner, D. (2013). *A frequency dictionary of contemporary American English: Word sketches, collocates and thematic lists*. Routledge. <https://doi.org/10.4324/9780203880883>
- Essam, B. A. (2017). Compiling a lexicon of pornography using web, WordNet and Frame Net to develop an individual pornographic index. *Sexuality & Culture*, 21(2), 534–548. <http://doi.org/10.1007/s12119-016-9403-9>.
- Hanson, R. C., Rose, E., & Little, Z. (2001). A comparative investigation of the semantic structure of language. *Cross-Cultural Research*, 35(3), 303–342. <http://doi.org/10.1177/106939710103500303>
- Holes, C. (2018). *Arabic Historical Dialectology: Linguistic and Sociolinguistic Approaches* (Vol. 30). Oxford University Press. <https://doi.org/10.1093/oso/9780198701378.001.0001>
- Ingham, B. (1994). *Najdi Arabic: Central Arabian* (Vol. 1). John Benjamins Publishing. <https://doi.org/10.1075/loall.1>
- Ismail, S. M. (2017). Translating dialect embedded in Najdi proverbs into English. *Dialectologia: Revista Electrónica*, 19, 93–107.
- Jorgensen, J. C. (1990). The psychological reality of word senses. *Journal of Psycholinguistic Research*, 19(3), 167–190. <https://doi.org/10.1007/BF01077415>
- Mattiello, E. (2009). *Difficulty of slang translation*. In A. Chantler & C. Dente (Eds.), *Translation practices: through language to culture* (Vol. 122). Rodopi.
- McCrae, J. P., Wood, I., & Hicks, A. (2017, June). *The Colloquial WordNet: Extending Princeton WordNet with Neologisms*. International Conference on Language, Data and Knowledge (pp. 194–202), Springer, Cham. https://doi.org/10.1007/978-3-319-59888-8_17
- Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, 3(4), 235–244. <https://doi.org/10.1093/ijl/3.4.235>
- Wilks, Y. (2008). *Machine translation: its scope and limits*. Springer Science & Business Media.

Copyrights

Copyright for this article is retained by the author, with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).