



رواق العربية

Forum for Arabic Linguistics

Arabic Linguistics Forum Conference

List of abstracts

30th-2nd July 2020

Held at
University of Leeds



UNIVERSITY OF LEEDS

Table of Contents

June 30 Study day on the Arabian Peninsula

Phonation in Semitic Languages (invited talk)	7
<i>Barry Heselwood (University of Leeds)</i>	7
The Way to the Rainy Mountains: Semantic Networks of Natural Metaphors in Najdi Poetry	8
<i>Letizia Cerqueglini (Tel Aviv University)</i>	8
Diminutive proper nouns in Peninsular Arabic	9
<i>Mark Shockley (University of North Dakota)</i>	9
The Phonetic Realisation of Emphatics in Harsusi	10
<i>Hammal Saleh Al Balushi (University of Leeds)</i>	10
Affricate variation in Emirati Arabic	11
<i>Marta Szreder (United Arab Emirates University)</i>	11
The m- Definite Article in Upper Faifi Arabic	12
<i>Abdullah Alfaifi (Indiana University Bloomington)</i>	12
Perceived Identities and Dialect Accommodation in Bahrain	13
<i>Navdeep Sokhey (Virginia Tech)</i>	13
Intonational variation in the Arabian Peninsula: a study on yes/no-questions	14
<i>Aljawaharah Alzamil (University of York) & Sam Hellmuth (University of York)</i>	14
Disagreement and (Im)politeness in Saudis' Twitter Communication	15
<i>Sarah Almutairi (University of Leeds)</i>	15
Forensic Authorship Analysis: examining a corpus of Najdi Arabic tweets	16
<i>Mashaal Al Amr (University of Leeds) & Eric Atwell (University of Leeds)</i>	16
Speech rate as a tool to detect the status of laryngeal contrasts in Saudi Arabic plosives.	17
<i>Latifa Al Kuwaiz (University of York)</i>	17
The use of q and g by Najdi female and male preachers in religious discourse	19
<i>Majedah Abdullah Alaiyed (Qassim University)</i>	19
Notes on majhūr and mahmūs in light of the Quran, the Quranic reading traditions and the grammatical tradition (invited talk)	20
<i>Marijn van Putten (Leiden University)</i>	20

July 1st Arabic Linguistics Forum

Owning the means of discursive production: A sociolinguistic analysis of class in a Palestinian borderzone (invited talk)	21
<i>Nancy Hawker (independent scholar)</i>	21
Variation in Vowel Phonology across Traditional Negev Arabic Tribal Varieties	23
<i>Letizia Cerqueglini (Tel Aviv University)</i>	23
The Central Mount Lebanon Lebanese (CMLL) Vowel Space: A Laboratory Phonology Investigation	24

<i>Georges Sakr (University of Edinburgh)</i>	24
Limbi Arabic: phonology and class in contemporary Egyptian pop culture	25
<i>Islam Youssef (University of South-Eastern Norway)</i>	25
Length and lexical stress in Palestinian Arabic /a/	26
<i>Niamh Kelly (American University of Beirut)</i>	26
Cross-dialectal variation in phonetic realisation of the emphatic contrast in Arabic fricatives	28
<i>Mahmoud Alsabhi (University of York), George Bailey (University of York) & Sam Hellmuth (University of York)</i>	28
Laryngeal cues and the phonological emphatic contrast in spoken Arabic dialects	30
<i>Latifa Alkuwaiz (University of York), Rana Almbark (University of York) & Sam Hellmuth (University of York)</i>	30
The French voiced uvular fricative in Hedi Zaiem’s speech: a pilot study	32
<i>Lamia Trifi (The University of Texas at Austin)</i>	32
Acoustic Analysis of Blocking Consonants and Emphasis Spread	33
<i>Laura Rose Faircloth (The University of Texas at Austin)</i>	33
Depalatalisation of /gj/ and /kj/ in the Ḥarbi dialect in Medina: patterns of variation and change	34
<i>Mohammad Alrohili (Tabuk University)</i>	34
Emphasis and the Temporal Organisation of Consonant Sequences in Arabic	35
<i>Khalid Alsubaie (University of Leeds)</i>	35
Dorsal Rhotics in Arabic: Evidence From qeltu-Arabic	36
<i>Uthman Aldahook (Durham University)</i>	36
The Role of Prosody in Disambiguation of Disjunctive Questions in Jordanian Arabic: Evidence from Production and Perception	37
<i>Mohammad Bani Younes (University of York) & Sam Hellmuth (University of York)</i>	37
On the Relationship between Morphology and Reading Accuracy among Children in Israel	39
<i>Duaa Abu Elhija Mahajna (Bar Ilan University) & Abeer Asli-Badarneh (Bar Ilan University)</i>	39
Competing patterns of possession: Evidence from loanwords in Palestinian Arabic	40
<i>Lior Laks (Bar Ilan University)</i>	40
Initiated usage of case markers in Standard Arabic text production	41
<i>Lior Laks (Bar Ilan University) & Elinor Saiegh-Haddad (Bar Ilan University)</i>	41
The Geographical Distribution of the Definite Article in West-Arabian Dialects	42
<i>Francesco Grande (University of Turin)</i>	42
Definiteness Effects and the Arabic existential constructions	43
<i>Maris Camilleri (University of Essex) & Louisa Sadler (University of Essex)</i>	43
The morphological complexity in Arabic spelling	45
<i>Iyad Issa (Community Development Authority in Dubai)</i>	45
“P” variation in the different perfect constructions across Arabic	46
<i>Maris Camilleri (University of Essex)</i>	46
Participial Structures within the Left Periphery: the Case of the Arabic Active Participle ..	48
<i>Mohamed Naji (Moulay Smail University)</i>	48

Lost and newfound features: what urban varieties can tell us about the history of agreement in Arabic	49
<i>Simone Bettega (University of Turin).....</i>	<i>49</i>
Grammatical Development of the North African Purposive	50
<i>Mike Turner (University of North Carolina Wilmington)</i>	<i>50</i>
Universal Dependencies for Written Informal Arabic.....	52
<i>Valeria Pettinari (Roma Tre University).....</i>	<i>52</i>
Aspects of the Phonology and Morphology of Peninsular Dialects of Arabic (invited talk) 53	
<i>Stuart Davis (Indiana University).....</i>	<i>53</i>

July 2nd Arabic Linguistics Forum

Language and Identity in the pre-Islamic Arabic of the Levant and North Arabia (invited talk)	54
<i>Ahmad Al-Jallad (Ohio State University).....</i>	<i>54</i>
Kalām al-‘arab. Sources, Problems and Methods	55
<i>Francesco Grande (University of Turin).....</i>	<i>55</i>
A historical scenario fro the distribution of Tanwīn Alif in Christian Palestinian Arabic texts	56
<i>Phillip W. Stokes (The University of Tennessee)</i>	<i>56</i>
Non-Standard Written Practices in Andalusi and Maghrebi Medieval Texts.....	57
<i>Estefania Valenzuela Mochon (The University of Texas at Austin).....</i>	<i>57</i>
The role of education and dialect contact in the realization of (ðʕ) and (dʕ) in a speech community of Palestinian refugees in Syria	59
<i>Ourooba Shetewi (Newcastle University).....</i>	<i>59</i>
The Semantics of Substrate Lexica in Colloquial Levantine Arabic: Dynamics of Survival and Naturalisation.....	61
<i>Mila Neishtadt (University of Cambridge).....</i>	<i>61</i>
A sketch of the Bedouin dialect of the Shahibaat (Egypt)	62
<i>Valentina Serreli (University of Bayreuth).....</i>	<i>62</i>
Toward a reassessment of Tunisian village dialects: the case of Chebba	63
<i>Luca D’Anna (University of Naples “L’Orientale”)</i>	<i>63</i>
Parallel grammars in one language: An analysis of Arabic loanwords in Saraiki	64
<i>Nasir A. Rizvi (Lasbela University), Shah Bibi (Lasbela University), Tooba Sahar (Lasbela University).....</i>	<i>64</i>
“Face” in the Male World of Tribal Reconciliation (atwa) in Jordan.....	65
<i>Abeer Malkawi (University of Brighton).....</i>	<i>65</i>
Pranks as a new eliciting technique for the study of lexical diffusion in Moroccan Arabic 66	
<i>Jacopo Falchetta (IREMAM)</i>	<i>66</i>
fuṣḥá, ‘āmmīyah, or both?: a new theoretical framework for contemporary Arabic writing	67
<i>Saussan Khalil (University of Cambridge).....</i>	<i>67</i>

Lexical changes in Arabic newspaper writings: A corpus-based comparison of 2018 and 1950 Arabic newspapers	68
<i>Sarah Alajlan (University of Leeds) & Claire Brierley (University of Leeds)</i>	<i>68</i>
Corpus-based Grammatical Investigations in Media Arabic	69
<i>Ashraf Abdou (University of Oxford).....</i>	<i>69</i>
Functions and Collocations of “ʔeeh” in Egyptian TV Drama: A Corpus-based Study.....	70
<i>Mona Azam (The American University in Cairo).....</i>	<i>70</i>
The Role of Multilingual Patterns and Functions on Branding and Self Branding: An Investigation of Practices by Saudi Users on Twitter.....	71
<i>Reem Al Madani (Cardiff University).....</i>	<i>71</i>
CEFR Classified Arabic Vocabulary List	72
<i>Nouran Khallaf (University of Leeds), Serge Sharoff (University of Leeds) &.....</i>	<i>72</i>
<i>Micheal Ingleby (University of Leeds).....</i>	<i>72</i>
Evaluation of Association Measures in Automatic Extraction of Collocations in Arabic.....	74
<i>Souhaila Messaoudi (University of Leeds), James Dickins (University of Leeds) & Claire Brierley (University of Leeds).....</i>	<i>74</i>
Facing Time: A study of the spatial representation of time in Classical vs. Modern Arabic	75
<i>Mariam Aboelezz (Lancaster University).....</i>	<i>75</i>
Needs Analysis and Course Design in Arabic for Academic Purposes: A Case Study at a Saudi Arabian University in Riyadh	76
<i>Soha Altayar (University of Edinburgh).....</i>	<i>76</i>
Social Media language resources for Arabic dialect research: an evaluation	77
<i>Eric Atwell (University of Leeds).....</i>	<i>77</i>
Cross-cultural Pragmatics: apology Strategies in Libyan Arabic	78
<i>Ahmed Elgadri (Lancaster University) & Bethan Davies (University of Leeds).....</i>	<i>78</i>



Voices of the Arabian Peninsula

University of Leeds

June 30th, 2020



UNIVERSITY OF LEEDS



Phonation in Semitic Languages (invited talk)

Barry Heselwood (University of Leeds)

There are two ways of considering the role of the larynx in phonation, each of which has different implications for phonatory classification. We can consider the larynx as a tone generator, or as an airflow regulator. The first is an acoustic perspective, the second an aerodynamic perspective. Taking the larynx as a tone generator, we can classify sounds by whether they have a laryngeal tone or not, i.e. whether the vocal folds are vibrating or not. Sounds with a tone can be classed as ‘voiced’, sounds without can be classed as ‘voiceless’. This perspective leads to the classification of sounds with a wide open glottis, and also those with a closed or narrowed glottis, as ‘voiceless’. However, if we regard the larynx as an airflow regulator, we can divide sounds according to whether air is allowed to flow freely through the glottis or not. Sounds with a wide open glottis can be classed as ‘voiceless’ and those with a constricted glottis, preventing free flow of air, as ‘not voiceless’; this includes voiced sounds but also sounds produced with a glottal closure or narrowing. These different perspectives have led to different definitions of ‘voiceless’ (see Esling & Harris, 2005: 350-351). Instead of ‘voiceless’ and ‘not voiceless’ we can use the terms ‘breathed’ and ‘unbreathed’ (also used in Garbell, 1958, and Blanc, 1978).

It is clear from the division of Arabic sounds into *majhūr* and *mahmūs* by Sibawayh and other grammarians of the Abbasid period that they took an airflow-regulating view (Gairdner, 1935/1978: 189; see also Heselwood & Maghrabi, 2015: 134-139), although they do not identify what it is that does the regulating. A question that arises from Sibawayh’s account is whether the airflow-regulating view may in fact be the better one to take as a default for Semitic languages in general. A good place to start is with Proto-Semitic (PS), and in particular with PS emphatic consonants.

That the emphatics in PS were ejectives is now generally accepted (see Kogan, 2011: 61, for a summary of the arguments), as they still are in Ethio-Semitic languages. For ejectives, the glottis has to be firmly closed in order to initiate the airstream, therefore there is no laryngeal tone and no transglottal airflow. Under the tone-generating view, ejectives are ‘voiceless’ (e.g. in Kogan, 2011: 61), while under the airflow-regulating view they are ‘unbreathed’. In purely phonological terms, there may be no advantage in taking one view rather than the other for PS. In this presentation, evidence will be considered from subsequent diachronic changes affecting emphatics, their allophonic variation, and from morphophonological processes targetting emphatics, in order to evaluate whether phonation in Semitic languages is best handled in a voiceless-voiced framework or a breathed-unbreathed framework.

The Way to the Rainy Mountains: Semantic Networks of Natural Metaphors in Najdi Poetry

Letizia Cerqueglini (Tel Aviv University)

This research focuses on how semantic networks are implemented in Najdi Arabic oral poetry (NAOP), exploring universal vs. culture-specific traits in network implementation strategies. I consider the ad-Duwāsir corpus of Bedouin Arabic Najdi poetry collected by Kurpershoek (1994–2002). In natural language processing, the semantic network model (SNM; Richens, 1956) represents semantic relations between concepts as a dynamic net. Modern SNMs calculate directions of semantic associations and predict the extension and limits of semantic networks' implementation. Interestingly, SNM prototypes served in ancient logic to improve mnemonic techniques (see Porphyry's *On Aristotle Categories*; Strange, 2014). NAOP largely resorts to mnemonic techniques and formulaic patterns in composition and transmission, employing complex networks of recursive metaphors that convey several connected concepts over many verses. Poetic language is ideal for this study, being a source for metaphor networks and their creative implementation (Herzog, 2012). First, I isolate and classify recursive metaphor networks, using Carston's model (2010) for relevance theory (Wilson & Sperber, 2004), which analyzes communicative intention in literary texts and everyday conversation, dividing metaphors into propositional and image-based types, both used in NAOP. Landscape elements and weather phenomena are the main sources of metaphorical transfer of feelings and emotions. My findings indicate the following: 1. Embodied experience of physical perceptions represents the cognitive basis of metaphorical communication (Lakoff, 1993; Croft & Cruse, 2004); 2. Embodying abstract experiences into physical phenomena streamlines communication (Evans & Green, 2006); 3. Landscape is one of the strongest factors shaping cognition and language (Mark et al., 2011). Nonetheless, the comparison of the NAOP metaphorical system with poetry from other Bedouin Arabic societies (Al-Ghadeer, 2009; Abu-Lughod, 1986; Bailey, 2002; Caton, 1990; Henkin, 2009; Holes & Abu-Athera, 2007; Husayn, 1926; Ritt-Benmimoun, 2009; Sergeant, 1951; Slyomovics, 2019; Socin, 1900) reveals that while they represent universal experiences, landscape and weather are subject to culture-relative interpretations (Orlove, 2003). Closely related, geographically contiguous cultures may select natural elements as metaphorical sources differently and map different meanings onto them. Finally, I analyze the trajectories of extension and the boundaries of NAOP metaphoric semantic networks. I report three fundamental networks: "inspiration," "sadness," and "nostalgia." Main findings concern: 1. "Partiality of metaphor" (cognitive networks connecting physical experiences are only partly exploited in metaphorical transfer; Lakoff & Johnson, 1999) and 2. Cultural specificities. In Najdi poetry, cognitively contiguous concepts such as "rain" and "storm" belong to distinct metaphor networks. "Joy/relief" and "inspiration," respectively, as in English "cold/chilly/ice" represent "lack of emotions," while "snow/floe/snowball" do not (Żołnowska, 2011). "Storm/strong wind/mountain top" are experientially contiguous concepts connected to "inspiration." Moreover, in NAOP, language specific dynamics constrain semantic metaphor networks' trajectories of expansion. Words are associated within the same metaphor network through assonance, rhyme, and other sound effects, and more often through morphological pattern than root. Language- and culture-specific preferences determine the expansion of metaphor networks, which are not exclusively and universally determined by physical experience. This study demonstrates the existence of language specific divergences between cognitive and linguistic metaphor networks.

Diminutive proper nouns in Peninsular Arabic

Mark Shockley (University of North Dakota)

This paper presents evidence that the syntactic diminutive has a special function in proper nouns in Peninsular Arabic. The most common diminutive forms in Arabic are $CuCayC$ and its quadrilateral forms $C1uC2ayC3iC4$ and $C1uC2ayC3iC4$. Less common diminutive forms include $C1aC2C2\bar{u}C3$, $C1aC2C3\bar{u}C3$, $C1aC2C3\bar{u}C4$, and, rarely, $C1aC2C2\bar{u}n$. In this paper, “diminutive” refers to this group of syntax patterns. The semantics of diminutive nouns in Arabic dialects is relatively well-investigated, but diminutive proper nouns have received little attention. In Iraqi Arabic, Masliyah (1997) noted numerous diminutive toponyms, anthroponyms, and wildlife names, but he found difficulty with these three categories (p. 77-78), all of which may be considered proper nouns. I propose that “properness” is a specific sense of the Arabic diminutive. This agrees with the diminutive’s sense of “exactness” noted in previous models, but the further argument for “properness” is a new contribution. For this study, 5454 toponyms and 2441 anthroponyms were gathered from Lorimer (1915), excluding Persian-speaking provinces. In addition, 428 plant names were gathered from Mandaville (2011). The following general observations were made:

1. The overall prevalence of diminutive forms is 19.58% for toponyms, 16.51% for anthroponyms, and 23.13% for plant names. Overall, out of 8323 Peninsular Arabic proper nouns, 1570 (18.86%) were diminutive. For comparison, secondary lexical sources on Gulf Arabic list 4909 common nouns, of which 211 are syntactically diminutive (4.30%).
2. Several diminutive nouns are used as generics in geographical names, especially hills and water sources; e.g. *ʕAʕainat Ḥasan* “the well of Hasan.” The same is observed in the *Gazetteer of Sudan* (U.S. Defense Mapping Agency, 1989).
3. In toponyms, the diminutive is applied repeatedly to names that denote largeness, prominence, and abundance (ʕ-z-m, ʔ-w-l, k-θ-r, etc.).
4. In many cases, the diminutive differentiates two similar proper nouns, regardless of size; e.g., *gabgūb* “lobster” is a diminutive of *gubgub* “crab”; *dbayy* “Dubai” is a diminutive of *diba* “Diba.”

To supplement these observations, a quantitative analysis was made of 5454 toponyms, which were coded using dichotomous variables and tested using chi-squared tests.

1. Areas typically associated with Bedouin-type Arabic dialects have a higher prevalence of diminutive place names ($p < .00001$). Kaye and Rosenhouse claimed that “diminutives are used mainly in Maghrebine and bedouin (and bedouinized) dialects” (1997, p. 284), but the claim was hitherto impressionistic and untested. There is also a negative link between permanent habitation and the diminutive form ($p < .00001$).
2. High prevalence of diminutive forms correlates with ownership ($p < .00001$), which suggests a link with linguistic properness.
3. I observed a strong positive statistical link between fresh water sources and the diminutive form ($p < .00001$), suggesting a connection to Bedouin naming practices.

These qualitative and quantitative findings are difficult to explain apart from a special use of diminutive syntax in proper nouns in the dialects spanning the eastern Arabian Peninsula. This also modifies current models of the semantics of the diminutive by adding “properness” as one of its possible senses.

The Phonetic Realisation of Emphatics in Harsusi

Hammal Saleh Al Balushi (University of Leeds)

Harsusi is one of the five Modern South Arabian (MSA) languages spoken in Oman from the total of six spoken in Oman and Yemen. The MSA branch belongs to the Western branch of Semitic language family tree. The MSA branch came to the attention of the European scholars around 1830s, however, Harsusi was not known in scholarly works until 1929. Since its discovery by Thomas Bertram in 1929, the scholarly work on Harsusi remains scarce. Today Harsusi is labelled as an endangered shifting language and the actual number of its speakers is unknown.

Similar to the other Semitic languages such as Arabic and Hebrew, Harsusi has a set of emphatic sounds which involves a secondary place of articulation. The earliest works on Harsusi (Johnstone, 1977 & Swiggers, 1981) mentioned that the emphatics were not pharyngealized as in Arabic, but rather glottalized as in other Ethiopic Semitic languages. This point was an interesting discovery and especially for historical linguists who assume the emphatics in the Proto Semitic language must have been glottalized (Faber, 1980; Rubin, 2019 & Huehnergard, 2019). Recent works in other MSA languages such as Mehri and Bathari have revealed that the emphatics were both pharyngealized and glottalized (Watson & Bellem, 2010, 2011 & Gasparini, 2018).

This paper presents the first results of an ongoing study in documenting the Harsusi language based on first-hand data gathered in the field after 43 years from the last work on Harsusi by Thomas Johnstone. The paper will explore in detail the occurrence of the two phonetic processes, pharyngealization and glottalization, in Harsusi. The paper will show that these both processes are involved in the realization of emphatic sounds in Harsusi as was found in some other MSA languages. It will further show that the articulatory realization of each emphatic sound is controlled by specific environments such as word initial, medial, and final positions. The data includes both elicited and natural speech recordings of 6 native Harsusi speakers. With regard to analysis, the paper will rely on the linguistic analysis program PRAAT to explore the acoustic cues related to Emphatic sounds. It will examine the oral burst, glottalic burst, the VOT, contextual creak, and F0 of the vowel following the Emphatic sound.

Affricate variation in Emirati Arabic

Marta Szreder (United Arab Emirates University)

Differences in affricate use and distribution are one of the defining features of cross-dialectal variation in Arabic. The affricates, [tʃ] and [dʒ], are derived from the Semitic *k [k] and *g [g] (Modern Standard Arabic [k] and [dʒ]), but exhibit significant variation across and within regions, varying between [k, ʃ, ts, tʃ] and [dʒ, ʒ, g, j] (Versteegh, 1997). Previous descriptions of this phenomenon in different Arabic dialects identified several conditioning factors of the variation, ranging from sociolinguistic variables (Dashti, 2018; Al Rojaie, 2013), to lexical (Hassan, 2017; Hoffiz, 1995) to vowel effects (Johnstone, 1963; Mustafawi, 2006). However, the variability of [dʒ] has traditionally been considered distinct from and more variable than [k] and usually not attributable to clear phonological effects (Feghali, 2007, Qafisheh, 1977).

In Emirati Arabic, [tʃ] often varies with MSA [k] and [j] often substitutes for MSA [dʒ]. The current study aims to document the patterns of this variation, which has not previously been experimentally studied. We combine evidence from three sources: corpus analysis, experimental elicitation data, and ultrasound tongue imaging, to investigate the conditioning factors of the variation, the within- and across-speaker stability of the variants, and their articulatory correlates.

The corpus part of the study was conducted using a spoken EA corpus (EMAC, Ntelitheos & Idrissi, 2017). The analyses reveal that the local variants [tʃ] and [j] commonly occur following the vowel [i:] and both are less likely to occur in the presence of the back vowel [u]. Furthermore, both appear to be facilitated by the presence of coronal segments – an effect not previously reported in other dialects.

The experimental data was collected in two picture-based elicitation experiments, targeting the segments in various phonological neighborhoods. The results of both experiments reveal a significant amount of both within- and across-speaker variability that is not lexically determined, as well as confirming the role of coronal segments.

The ultrasound data was collected from participants during the elicitation experiments. The analysis of tongue surface imaging using smoothing-spline analysis of variance (SSANOVA) demonstrates a two-way contrast for [k ~ tʃ] and a three-way distinction for [dʒ ~ ʒ ~ j] with the plosive variant being categorically distinctive and the fricative variant appearing to be gradient. Furthermore, the results suggest that the difference between the variants is primarily in constriction degree.

Combining the data from the above three sources, we conclude that affricate variation in EA is partially lexically conditioned, and partially sensitive to the phonological context, including front vowels and coronal consonants. However, while the two local variants [tʃ] and [j] pattern together in this respect, they are different in the amount of exhibited within- and across-speaker variability, as well as in the distinctiveness of their articulatory correlates. We propose that both [k~tʃ] and [dʒ~j] variation are similar in nature and involve a change in constriction degree, but may be at different stages of development.

The m- Definite Article in Upper Faifi Arabic

Abdullah Alfaifi (Indiana University Bloomington)

In Modern Standard Arabic (MSA) and most Arabic dialects, the definite article is l- (or al-, il-, lə-), which surfaces in various assimilatory forms which reflect a following coronal consonant, or alternatively, true coronal geminated realizations (Zaborski, 2006; Heselwood & Watson, 2015). Some Arabic dialects exhibit other forms of the definite article including b-, n-, and m- (Rabin, 1951; Alqahtani, 2015; Heselwood & Watson, 2015; Behnstedt, 2016). This paper focuses only on the m- realization of the definite article in Upper Faifi Arabic (UFA), a dialect spoken in Faifa Mountains in the southwest of Saudi Arabia. The m- definite article in this dialect attaches to most nouns, including some proper nouns such as [jaḥja m-saalim] for ‘Yahya Salem’ (Alfaife, 2018), and adjectives [mbajt imśabiir] ‘the big house’. While most Arabic varieties have the definite article /l-/, it is almost exclusively prefixal /m-/ in UFA. As found in other nearby dialects, UFA possesses both l- and m- as definite articles, but there appear to be certain semantic and syntactic factors that determine the occurrence of either l- or m-.

The obligatory use of the m- definite article for most words makes it different from other varieties of southwestern Saudi Arabic (SSA) where there is optional variation on the same word between the use of the l- definite article and the m- definite article. (Alqahtani 2015; Alahmari 2015). Nonetheless, in UFA, for semantic reasons, the m- definite article is blocked in words with religious significance, the names of countries, the days of the week, months, and names of nonlocal cities. For example, the MSA word [ʔal-madiina], referring to ‘Al-Madīnah AlMunawwarah’, is realized similarly in UFA, but when the same word is used to mean ‘the city’, the UFA realization is [m-madiina]. Further, names of days of the week take the l- definite article as in [jawm il-xamiis] ‘Thursday’. While the names of the days appear in the Idafa construction, the Idafa normally takes m- as in [baab im-bajt] ‘door of the house’.

Further, as observed by Alfaifi & Behnstedt (2010: 60), the m- definite article is typically not used with a demonstrative in Faifi Arabic. While this observation is accurate for a pre-posed demonstrative as in (a), the m- definite article can be used with a post-posed definite article as in [m-marah ti] ‘this woman’, which has the same meaning as (a). Moreover, if the prefixal [m-] definite article were to be used with a demonstrative as in (b), then the expression would be understood as a complete sentence rather than as a noun phrase. This, thus, distinguishes between an inherent ambiguity that can be found in other dialects with respect to [til-marah] as either ‘this woman’ or ‘this is the woman’. This paper will examine these properties as well as other properties of the m- definite article in UFA.

- | | | | | | | | |
|----|------|-----|---------|----------------------------|------|-----|--------|
| a. | [ti | l- | marah] | <i>‘this woman’</i> | *[ti | m- | marah] |
| | DEM | DEF | woman | | DEM | DEF | woman |
| b. | [ti | m- | marah] | <i>‘this is the woman’</i> | *[ti | l- | marah] |
| | DEM | DEF | woman | | DEM | DEF | woman |

Perceived Identities and Dialect Accommodation in Bahrain

Navdeep Sokhey (Virginia Tech)

Third-wave variationist theories posit that linguistic features linked to ideological meanings have indexical powers, allowing their meanings to be continually re-negotiated with every interaction (Eckert 2012, Podesva 2004). The assignment of certain social values to linguistic features is thus often interaction-specific (Sharma and Rampton 2015). To date, no Arabic sociolinguistic study explores variation based on the perceived identities of the interactants using interactional analysis as a method. This paper examines rounding of the long vowel alif [aa] → [ʊ] in the speech of 30 Bahraini university-aged speakers when interacting with other Bahraini interlocutors of various perceived communal backgrounds and gender. Previously, vowel-rounding was noted to be a regional feature of women from the old Bahraini city of Muharraq (Holes 1987). It was observed to be a feature of the socially dominant dialect of Sunni Bahraini sand to occur in the speech of some school children of both Sunni and Shi'i backgrounds (Al-Qouz 2009). However, the complex usage of this feature in real-time interactions is left uninvestigated. Using interactional analysis, this study offers a new methodology to Arabic dialectology for approaching real-time dialect accommodation. The methodology employed can be modified and adapted to study other speech communities. The present study traces the construction of social meanings surrounding vowel-rounding alongside communal identity using a linguistic accommodation experiment. It examines how Arabic speakers in Bahrain today vary their dialect choices and explores their use of vowel-rounding based on their own social background as well as the perceived identities of their interactants. The experimental task consisted of a simulated Q&A interview using pre-recorded questions produced by 4 native speakers in either the Sunni or Shi'i dialect, each question containing either a rounded or unrounded vowel occurring in stressed syllables, surrounded by voiced or voiceless consonants. Participants answered by restating each question in its entirety, revealing their selection of dialectal features indexing a particular community alongside a rounded or unrounded vowel. A total of 32 questions generated over 2,500 tokens from the 30 participants. The data were analyzed statistically and the results were interpreted based on a year of ethnographic work in Bahrain. Using a mixed effects model, the degree of vowel-rounding was acoustically measured by extracting the F2 values. The continuous, dependent variable Rounding was then analyzed against the independent variables Sex (coded as Female/Male), Perceived Communal Identity (Sunni/Shi'i), as well as the "actual" Communal Identity of both the participants and their interactants. The results indicate that actual and perceived gender and communal background of the interactants significantly affected both the dialect choice and use of rounding by participants, in that rounding is exhibited most strongly in the speech of the Shi'i male participants, and increased when interacting with other men and other "in-group," Shi'i women. This not only demonstrates an innovative use of rounding to index masculinity – a quality not stereotypically associated with this sound, but further suggests that a reclamation of rounding as a nationalistic marker of a "Bahraini" identity transcending communal affiliation is in progress.

Intonational variation in the Arabian Peninsula: a study on yes/no-questions

Aljawaharah Alzamil (University of York) & Sam Hellmuth (University of York)

Prior studies on Arabic intonation have demonstrated that cross-dialectal studies are a fruitful area due to a high degree of variability across dialects (ElZarka, 2017), and, further, that the nuclear contour of yes/no-questions (YNQs) in particular can be a useful discriminating feature (Hellmuth, 2018). This study presents qualitative analysis of the prosodic contours observed in YNQs in three Saudi Arabic dialects (Najdi, Hijazi, and Jizani), with comparison to data in the Kuwaiti Bedouin dialect, to explore the possibility that intonation contours can be used to differentiate geographically adjacent Arabic dialects. More broadly the study seeks to contribute to the description of the intonation patterns of these dialects and to identify potential examples of sociolinguistic and/or dialectal variation for further investigation. The data come from a parallel corpus elicited using the same methods as in the Intonational Variation in Arabic corpus (Hellmuth & Almbark, 2017), which is in turn modelled on the techniques used in the Intonational Variation in English project (Grabe, 2004). Data for the pilot study were collected with 10 speakers of Saudi dialects temporarily resident in the UK (Najdi: 2F; Hijazi: 2M/2F; Jizani: 2M/2F), for comparison to IVAr corpus data from four Kuwaiti Bedouin (KB) speakers [2M/2F]. Data for the main study were collected with 72 speakers of Saudi dialects citizens in Saudi Arabia (Najdi: 12F/12M; Hijazi: 12F/12M; Jizani: 12F/12M). We analysed read speech YNQs elicited in a scripted dialogue (N=497) and semi-spontaneous YNQs extracted from map tasks (N= 520). The results of the pilot study show that YNQs are typically realised with one of the two contours previously observed in Arabic: a rise (Hijazi, Najdi and female KB) versus a rise-fall (Jizani and male KB). The patterns observed in pilot study read speech (as illustrated in Fig.1) were also found in pilot study semi-spontaneous speech and main data read speech. Finding both contours in adjacent dialects in the same geographical area suggests there is a boundary for the isogloss between the rise and rise-fall in this region; the final main data results will help to determine where the boundary is and whether it is moving. Differences in contours typically used by male/female Kuwaiti Bedouin speakers suggest that the shape of the yes/no-question contour would be a useful variable for inclusion in a variationist sociolinguistic study.

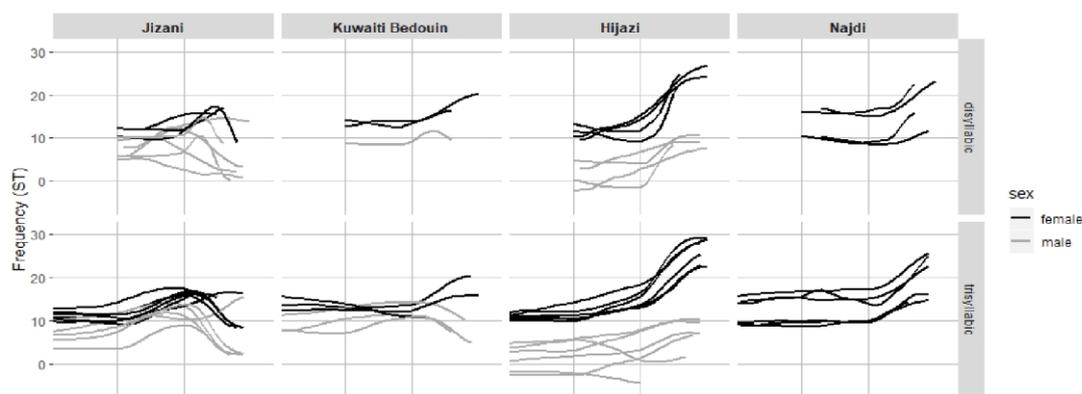


Fig 1: Time-normalised smoothed F0 in polysyllabic utterance-final words in read-speech yes/no questions in pilot study (N=59), by dialect and sex; each vertical column in the grid represents one syllable

Disagreement and (Im)politeness in Saudis' Twitter Communication

Sarah Almutairi (University of Leeds)

The paper discusses disagreement and (im)politeness in Twitter communication, focusing on social and political trending hashtags in Saudi Arabia during 2017-2018. This initial stage of the research consists of a corpus investigation of disagreement strategies used by Saudis and investigates how disagreements are linguistically mitigated (i.e. weakened) or aggravated (i.e. strengthened). The analysis identified 15 distinct disagreement strategies, and the bulk of the tweets (49%) showed multiple strategies being used in combination.

The analysis further reveals that classifying disagreement strategies as polite, impolite, politic, etc. based on Locher & Watts' (2005) relational work model is not a straightforward process. It is true that some of the disagreement strategies are clearly negatively marked hence classified as impolite/inappropriate like verbal attack and verbal irony, yet the majority are unmarked hence not easily classified. For instance, a disagreement in the form of criticism can be either impolite/inappropriate, polite/appropriate, or politic depending on the existence of another negatively marked strategy as well as the existence and the impact of mitigation or aggravation devices used with the disagreement strategies.

Moreover, it is important to highlight that mitigation in some cases can be ineffective or even impolite, this use of mitigation is mostly found in sarcastically mitigated disagreement. Accordingly, the relational work model was modified to include the sarcastically mitigated disagreements, the model consists of 4 major types of disagreements but based on the markedness there are 3 general categories. The first category is the positively marked disagreements, which covers mitigated disagreements. The second category is unmarked, which includes disagreement instances that are neither mitigated nor aggravated. Lastly, the negatively marked category encompasses both aggravated disagreement and sarcastically mitigated disagreement. The overall results show that mitigated disagreements have a low frequency in the corpus at 10%, unmarked disagreements at 43% and aggravated disagreements have a higher frequency at 47%.

Forensic Authorship Analysis: examining a corpus of Najdi Arabic tweets

Masha'el Al Amr (University of Leeds) & Eric Atwell (University of Leeds)

Najdi Arabic has been the subject of research in phonetics (Badawi, 2012), morphology, syntax (Al Sweel, 1992; Binturki, 2015), sociolinguistics (Ingham, 1995), and Computer-Mediated Communication (Alothman, 2012). Also, several studies in computational linguistics investigated Saudi Arabic (Alruily, 2012; Althenayan and Menai, 2014; Al-Tuwairesh et al., 2015, 2018; Assiri et al., 2016), but there is much to explore in the forensic linguistics domain (Mansour, 2013). Bucholtz and Hall (2005) claim in their identity approach that interaction can reveal and generate an online identity represented in the user's idiolectal style. This claim is further supported by forensic linguists Hardaker (2010), McMenemy (2002), and Turell (2010) who describe idiolectal style as a product of socio-cultural elements and social experiences. This paper examines these concepts in a corpus of thirteen authors, six males and seven females, from the central region of Najd in Saudi Arabia. The corpus of 748,348 words is composed of six male and seven female participants' tweets with a reference corpus representing the relevant Najdi population (Larner, 2014; Heydon, 2018). The methodology combines computational and stylistics approaches using Data Miner to collect data and Wordsmith to extract frequency rates along with a critical discourse analysis of how the linguistic features construct the authors' online identity. The analysis explores the individual idiolectal style of each author that emerges through interaction in their use of vocabulary (both dialectal Najdi and Modern Standard Arabic). The principles of the identity approach help to clarify how these authors use interrogatives, negatives, and deictic expressions revealing how they position themselves in the community. Dialect specific items such as the interrogative pronouns *what* / *وش* and *why* / *وزاه وشوله* and the negative *not* / *منب / مب* show unique patterns of use across the authors' subcorpora. The unique usage of these Najdi-specific items can help not only to identify a group of authors who use the same dialect but also to isolate each unique identity.

Speech rate as a tool to detect the status of laryngeal contrasts in Saudi Arabic plosives

Latifa Al Kuwaiz (University of York)

Speaking rate typically has asymmetric effects on voice onset time (VOT) crosslinguistically: as speech rate decreases, values of long-lag and voicing-lead both increase, but short-lag VOT is not affected. Phonologists have thus proposed that in two-way laryngeal contrast languages (e.g. English: [sg] vs. [Ø]) only one “active” side of the phonological contrast is affected by speech rate (Summerfield, 1975). However, recent results from Swedish (Beckman, Helgason, McMurray, & Ringen, 2011) and Qatari Arabic (Kulikov, 2015) show rate effects which support overspecification of both [voice] and [sg] as active features in a two-way phonological contrast.

Arabic dialects vary in the mapping of VOT values to laryngeal contrasts (Bellem, 2007), providing rich ground to explore this overspecification analysis. This paper uses speech-rate to determine the active feature(s) in two dialects of the Arabian Peninsula: one is predicted to have a two-way VOT distinction (Hijazi) and the other a three-way distinction (Najdi). The paper also examines the nature of the laryngeal contrast in the Saudi Arabian Hijazi variety which, to our knowledge, has never been examined before. Three dependent variables (VOT/closure duration/% of voicing) are measured in production data for eight plosives: plain /t, d, k, g, q/ and emphatic /t/ and /d/¹, placed in word-initial/medial positions, followed by three long and three short vowels. The stimuli are real/nonsense words presented to the subjects in Arabic orthography. Data was collected with 18 speakers per dialect (Hijazi 10F/8M; Najdi 8F/10M) who were asked to read the stimuli at two different speech-rates. Manual labelling in Praat of the plosive release (burst), onset of voicing and % voicing during closure is ongoing, and data will be explored with linear mixed-effect models (LMM) in R.

Results of pilot study data using similar design (1F-Hijazi/ 1F-Najdi) showed successful speech-rate effects using word-duration as proxy for rate (*figure.1*). In the pilot study, speech-rate showed dialect-specific effects on VOT. In both dialects, voicing-lead in voiced plosives shows a clear speech-rate effect, but the effect of speech-rate on long-lag VOT in /t/ is smaller (i.e. more overlapping range of values at slow/fast rate) in Hijazi than in Najdi. Although an interim result only, this trend is in line with our predictions, and points to potential sound change in progress in Hijazi. In the paper, we will present results of ongoing analysis of the full dataset along with discussion of the implication of observed speech-rate effects for our understanding of the underlying representation of the laryngeal contrast in the Saudi Arabic plosive system.

¹ A subscript dot rather than IPA diacritics is used to distinguish emphatics from plain counterparts. This is to avoid assumptions about phonetic realization. For the same reason, ‘emphatic’ rather than e.g. ‘pharyngealized’ is used.

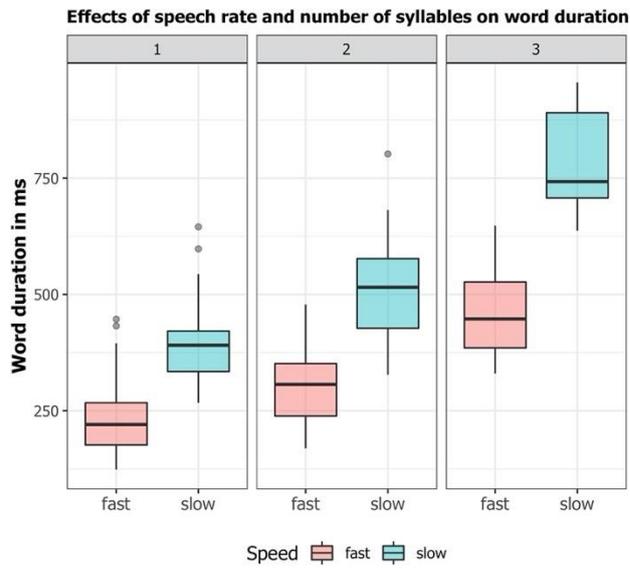


Figure.1

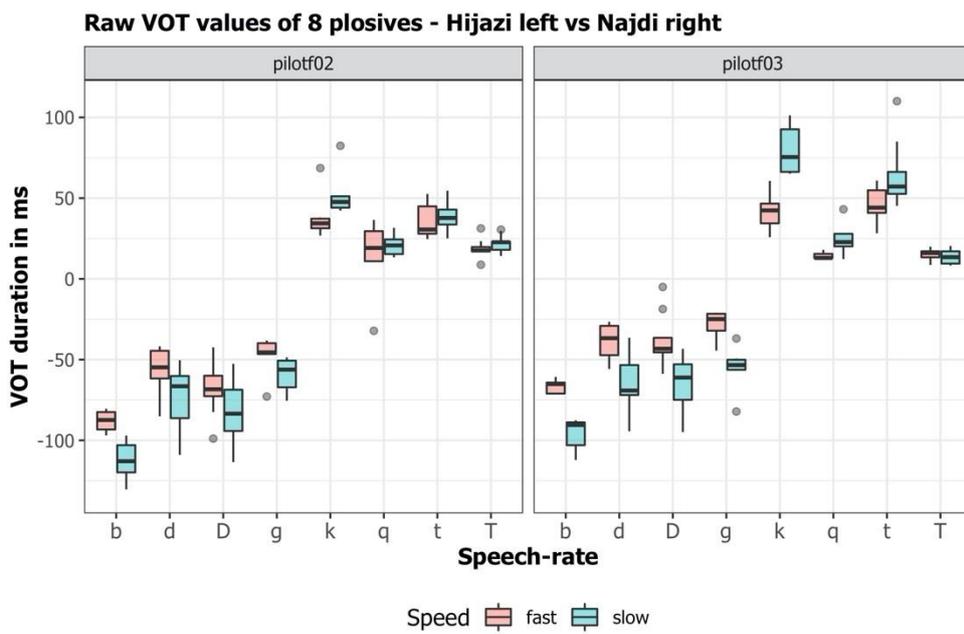


Figure.2

The use of q and g by Najdi female and male preachers in religious discourse

Majedah Abdullah Alaiyed (Qassim University)

The phoneme q is a voiceless sound produced “from further back in the mouth – from the uvula, to be exact” (Holes 1994:10). Thus, it is a voiceless uvular stop. For several reasons, including “dialect mixing and processes of Koineization” (Behnstedt 2006:596), the uvular stop has undergone a number of changes within and across various speech communities. For example, according to Ingham (1995), Najdi Arabic (NA) differs from Standard Arabic (SA) in its sound inventory. Amongst the new Najdi Arabic sounds is the voiced velar g, which is a result of the fronting of the uvular q. In addition, two new units, the voiced affricate ġ [dz] and the voiceless affricate c [ts], characterize the Najdi variety and have resulted from the fronting of g from Classical Arabic q and Classical Arabic k, respectively. As a result, Najdi Arabic now has an opposition between g/ġ and k/c. This opposition is limited in scope because in most cases the fronted varieties occur in front vowel environments (Ingham 1995).

This paper discusses a study on diglossic intra-sentential code-switching between Standard Arabic and Najdi Arabic in religious speeches by six Saudi preachers: three males and three females. The study focused on the alternation between the SA sound q and the NA g. It also focused on the difference between the males and females in their use of the two sounds.

Previous studies of religious speeches have focused only on male preachers, while previous studies of Saudi Arabic have disproportionately investigated the speech of men. Given the gender segregation of this speech context, and the sociolinguistic studies of Arabic showing some gender variation (Walters, 1991; Daher, 1998, 1999; Al-Wer, 1999), there is a question over whether there may be gender variation in code-switching in religious speeches in segregated speech contexts. By analysing the code-switching of both male and female preachers, the study makes an original contribution by demonstrating that there is a tendency by female speakers to use the Najdi sound g less than the male preachers do. Moreover, both genders show a preference for using the SA q more than the NA g.

Notes on *majhūr* and *mahmūs* in light of the Quran, the Quranic reading traditions and the grammatical tradition (invited talk)

Marijn van Putten (Leiden University)

The Sībawayh's categories of *majhūr* and *mahmūs* have often perplexed researchers. This has led to suggestions that the *q*, *ṭ* were voiced, and Sībawayh's description of the *hamzah* as *majhūr* was simply mistaken. But recently a renewed appreciation of Sībawayh's precision has developed. Due to the revolutionary work by Heselwood, Watson and Maghrabi our understanding of what the terms *majhūr* and *mahmūs* entail has shifted radically. It is now clear that the presence or absence of turbulent airflow was the main distinguishing factor that led to this binary classification.

Despite Sībawayh's development of this binary distinction, he does not make use of this distinction to explain any phonetic phenomena that may be related to it in his *kitāb*. However, his insights on the distinction between *majhūr* and *mahmūs* are of vital importance to our understanding of Quranic rhyme. While often described as 'chaotic', the Quranic rhyme in fact follows a very precise system of rhyme once one takes into account the distinction between *majhūr* and *mahmūs* (incidentally Quranic rhyme bears a striking similarity to Old Irish rhyme). In the Quran we encounter three types of rhyme:

1. Resonant rhyme (r, l, m, n)
2. *Majhūr* rhyme (the *majhūr* consonants including the resonants)
3. *Mahmūs* rhyme (which is almost entirely consists of monorhymes)

However, from this study of the rhyme of the Quran, an interesting anomaly shows up: for Sībawayh there is one emphatic consonant that does not belong to the *majhūr* class, namely, the *ṣād*. In the Quran, however, *ṣād* rhymes as if it is a *majhūr* consonant. As we will show, the canonical Quranic reading tradition of the Kufan Ḥamzah (d. 156 AH/773 CE) appears to have retained a marginal memory of the original *majhūr* articulation of the *ṣād*.

Finally, we will explore the possibility of the existence of pausal glottalisation in Quranic Arabic and Classical Arabic. In both Yemeni Arabic and the Modern South Arabian languages, pausal glottalisation appears to be specific to the *majhūr* consonants. While no direct references to glottalisation of these consonants can be found in the Arabic grammatical tradition, there is some evidence for pausal glottalisation of final vowels, which opens the door to exploring possible parallels between the pausal effects in southern Arabia and those found in Quranic Arabic.

Owning the means of discursive production: A sociolinguistic analysis of class in a Palestinian borderzone (invited talk)

Nancy Hawker (independent scholar)

Only sociolinguistics can explain an occurrence such as this one:



A Yiddish poster for the Arab-led Joint List القائمة المشتركة party ahead of the 2 March 2020 elections for the Israeli parliament, translating as: “Your voice against compulsory conscription”. Yiddish is a Germanic language written in Hebrew script and is used by ultra-orthodox Jews in Israel and elsewhere. © The Joint List (in Yiddish, פאַרײניקטע רשימה)

Sociolinguistics makes a big promise: using its methods of analysis reveals the traces that past and ongoing social processes leave in languages. For example, migration is a social process that leads to language contact. This contact resulted in, say, historical Coptic influence on negation in Arabic (Lucas & Lash 2010). Furthermore, these traces can then be mobilised in discourse to index where speakers stand in relation to social processes (Agha 2011; Jaffé 2016; Silverstein 2003). These stances can also be analysed with the help of sociolinguistics. Such is the case, for instance, of young men in Amman modifying their realisation of Qaf depending on their interlocutor (Al-Wer & Herin 2011).

The social processes in question are usually dissected along broad sociological categories: gender, age, ethnicity and religion or race, geographical location of early education or residence, socioeconomic class, etc, which are constitutive of identities. Beyond the dissection, when linguistic traces, their indexing, and stances are synthesised, specific ways of speaking can themselves create sociological categories (Irvine & Gal 2000). This has been demonstrated time and again in relation to nationhood. Languages frequently define nations, in contested and incomplete ways (Cameron 2006). And these linguistic national formations can also be analysed sociolinguistically. This can be seen, for instance, in the recent political

recognition of Tamazight (Berber) in North African countries, and the debate over which variety to standardise (Soulaïmani 2016).

Yes, sociolinguistics can explain almost everything – probably not astrophysics – on various scales and from different perspectives. The big promise of this lecture is the application of sociolinguistics to the case of Arabic in arguing the case for highlighting class as a primary sociolinguistic category. Drawing on evidence gathered ethnographically over 15 years in a relatively small area southwest of Jerusalem, the lecture will present excerpts of speech from Palestinian manual labourers, mall shoppers, olive growers, and politicians, chosen from a total corpus of 135 speakers (Hawker 2013, 2019, 2020). These excerpts will be analysed using critical discourse analysis (Wodak & Meyer 2015) and systemic functional linguistics (Young & Harrison 2004).

The manual labourers, just as the politicians, use Hebrew loanwords in their Arabic speech to display their knowledge in specific subjects. Some of the mall shoppers mix in Hebrew discourse markers for a ‘cool’ slang that matches their consumerism. Olive growers maintain inherited agricultural knowledge with inherited local terminology. And among all of them are jokers, perhaps more accustomed than others to holding the floor and engaging interlocutors, who also use Hebrew for ironic humour. The difference between the various groups of speakers lies in degrees of confidence in their discursive authority (Gal & Woolard 2014). They are different in degrees of privilege and precarity (Butler 2009) which inform their confidence or lack thereof.

There are speakers who confidently exploit their linguistic repertoires, as they would material resources (Heller 2003). To paraphrase Karl Marx, some privileged speakers are producers of discourse: they own their languages as a means for their (self-)production. This conceptualisation offers a solution to the tension between structure and agency that suffuses sociolinguistic theory. In a conflictual borderzone, a ‘viscous space’ (Parizot 2018), sociological categories such as ethnicity and gender form structures that appear impervious to agency, or even bind their subjects in positions that make them vulnerable to precarity. In this context, class retains a potential for agency that makes the occurrence of discourse possible. The Arab creators of the Yiddish poster own their repertoires as means to produce discursive authority. These themes are brought into sharper, and more paradoxical, focus in the contested and incomplete nation-forming projects in Israel and Palestine than elsewhere. Nevertheless, they illustrate speakers’ agency behind representations of complex identities that sociolinguistics can analyse the world over.

Variation in Vowel Phonology across Traditional Negev Arabic Tribal Varieties

Letizia Cerqueglini (Tel Aviv University)

First described as one language (Blanc, 1970), Negev Arabic (NA; southern Israel) was considered a koine of tribal varieties with various origins (Henkin, 2010) that converged around the Negev from the seventeenth century (Bailey, 1985). Today, NA is defined as a cluster of Northwestern Hījāzi Bedouin dialects spoken across Sinai, the Negev, and Jordan (De Jong, 2000; Palva, 1984, 1991, 2008). Different NA dialects are spoken in the main Negev Bedouin confederations— ʕAzāzmih, Gdērāt, Ẓullām, Tarābīn, Tiyāha, Maʕāniyyīn—and generally thought to be differentiated by lexical variants and some features of verbal and nominal morphology (Shawarbah, 2007, 2012). I compare the phonological systems of Traditional NA varieties (TNA) spoken by elders over age 70. I show that TNA dialects have different vowel phonemes, stress rules, and stress changes in pausa. In particular, TNA varieties differ in type and degree of imāla (‘inclination’). Imāla indicates the rising and fronting of /a/ towards /i/ and /ā/ towards /ī/ in various contexts and under different conditions. Dialectologists use this phenomenon, described by early Arab grammarians (Derenbourg, 1881–9), for classificatory purposes (Blanc, 1964; Levin, 2007; Owens, 1993; Shawarbah, 2011). Across Arabic dialects, imāla reaches different degrees (æ, ε, e, i) and occurs under different conditions in all contexts, in certain vocalic and/or consonantal environments (Levin, 2007), and in accordance with morphological category (Blanc, 1964). According to my 480,000-word corpus of TNA narrative, poetry, and daily speech in different tribal dialects, TNA-Gdērāt systematically shows strong, unconditioned imāla. Close to /y/, [e] and [ē] appear. In other cases, the imāla degree is lower, while ā is not a phoneme and [ē] is not phonologically distinguished from /ī/. TNA-Tiyāha imāla is conditioned by the vocalic environment and is lower than imāla in Gdērāt. In general, TNA-ʕAzāzmih imāla follows the morphological category, never occurring in plural adjectives. Yet the ʕAzāzmih confederation is not phonologically unitary; some groups (e.g., TNA-Tarābīn and TNA-Ẓullām) have imāla only in final /a/. TNA-Maʕāniyyīn has no imāla. Imāla produces a series of implicational effects in the phonological system of the short vowels. TNA-Gdērāt has /æ/ instead of the TNA common /a/, with [a] close to pharyngeal and laryngeal consonants. Epenthetic vowels are used in all TNA varieties, only between -CC clusters at the ends of words. In TNA-Maʕāniyyīn, vowel harmony affects the timbre of the epenthetic vowels; these are generally longer than in other varieties. Surprisingly, in TNA-Maʕāniyyīn, epenthetic vowels are stressed in pausa (*gámr→gámar→gamár#). A further phonological feature of TNA-Maʕāniyyīn is a special outcome of Gaháwah syndrome (Blanc, 1970), which is found in all North Arabian Bedouin dialects and consists of *(C)aXC(V)→(C)aXa(C)(V), with X = ġ, ḥ, ḥ, ʕ, h. An epenthetic vowel of low-central quality is inserted to create a new syllable. In all TNA varieties, this syllable is stressed and the first vowel often falls. In TNA-Maʕāniyyīn, the first vowel always remains stressed. The systematic description of the phonological systems of TNA tribal dialects reveals different linguistic histories and contributes to the internal classification of the North-western Arabian dialectal continuum.

The Central Mount Lebanon Lebanese (CMLL) Vowel Space: A Laboratory Phonology Investigation

Georges Sakr (University of Edinburgh)

Most references speak of the Arabic vowel space (eg. Al-Ani, 1970) as a three-vowel system. Starting with a phonetic analysis of 4355 vowel tokens of acoustic CVC data I collected from 19 native speakers of Central Mount Lebanon Lebanese (henceforth CMLL) in 2018-2019, I show that the underlying (or contrastive) vocalic inventories of CMLL has seven distinct monophthongal vowel qualities, and I situate this result within the established typology of vowel inventories in terms of symmetry and inventory size.

The vowels were hypothesised from the literature and through native intuition. Participants were asked to complete sentences that elicited words using each of the vowels, in Lebanese Chat Orthography to reduce influence from Standard Arabic. The words were suggested by their consonants to avoid lexically similar forms (eg. *buw:e:be* ‘gate’ for *be:b* ‘door’). I isolate the most economical set of contrastive features from my set of contrastive segments, and these features suggests that the most sensible division of the underlying vocalic inventory (monophthongs and diphthongs included) of CMLL is into four distinct systems (see figure 1): short monophthongs, long monophthongs, long monophthongs that vary in production, and diphthongs.

Two of these systems do not fit typological expectations in that they feature a front-back contrast in low vowels. While this is regularly attested in six-quality systems, of which the CMLL long monophthong system is one, it is particularly peculiar in the CMLL three-quality short monophthong system, in which the high vowels have undergone a merger (as suggested by e.g. Haddad, 1984; Watson, 2002). Another important divergence from typological expectation is the fact that this merger of short high vowels goes against Maddieson’s (1984) assertion that, for languages with multiple vowel systems, vowels in one system can usually be matched with vowels of similar qualities in other systems, such that ‘the overall number of vowel phonemes is greater than the number of different vowel qualities’.

Central Mount Lebanon Lebanese vowel qualities follow a 6L1 structural configuration for vocalic inventories (see Becker-Kristal, 2010), which makes it comparable to Babine, Estonian, Gujarati, Dutch, Occitan, Chichimeca, Albanian and Wolof. A comparison with Maddieson (1984) shows that, since the CMLL vocalic inventory I proposed consists of 9 non-varying monophthongal phonemes, it has more monophthongal vocalic phonemes than 63.8%, and less than 27.4%, of the world’s languages. These phonemes cover 7 vowel qualities, which means that CMLL has more monophthongal vocalic qualities than 63.7%, and less than 21.5%, of the world’s languages.

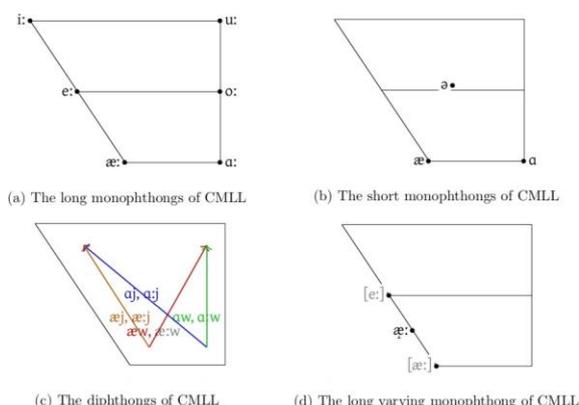


Figure 1: The Vocalic Systems of Central Mount Lebanon Lebanese

Limbi Arabic: phonology and class in contemporary Egyptian pop culture

Islam Youssef (University of South-Eastern Norway)

This paper investigates the phonological patternings in the speech of il-Limbi, an immensely popular character in Egyptian comedy played by actor Muḥammad Sa‘d; and it stands therefore at a crossroads between cultural studies and linguistics. Il-Limbi represents the urban working classes, and his speech often mocks social conventions through ludicrous parody of educated speech (Hamam 2012). Masquerading as socially superior personas, his speech highlights the diglossic situation in Egypt as well as the pretentious use of English into the elite register.

My examination of il-Limbi’s pronunciation in four films reveals a number of systematic patterns in both consonants and vowels, which construct a unique code that I dubbed “Limbi Arabic”. An example of consonantal patterns is regressive voicing of sibilants, whereby the voiceless alveolar /s, ʃ, š/ become voiced /z, ž, ž/, when followed by a voiced obstruent (see Watson 2002:245), as well as in inter-vocalic and coda positions (1a). Another example is mutation (Woidich 2006:15), whereby the consonant’s place or manner of articulation is altered; the changes show a preference for nasal over oral outputs and for alveolar over non-alveolar consonants (1b). Other trademark features include metathesis, as in *ru ‘b-i gnē* for *rub ‘-ignē* ‘a quarter of a pound’, and guttural lenition, a change from voiceless pharyngeal or laryngeal /ʔ, ḥ, h/ in pre-vocalic position to the voiced pharyngeal /ʕ/ (1c).

(1) Limbi Arabic	Underlying	Gloss
a. <i>nilmiz burg</i>	<i>nilmis burg</i>	‘we touch a tower’
<i>illi biyžīl</i>	<i>illi biyšīl</i>	‘that carries’
<i>il-mikrubāž</i>	<i>il-mikrubāš</i>	‘the microbus’
b. <i>ğaşm-i ‘anni</i>	<i>ğaşb-i ‘anni</i>	‘against my will’
<i>fī-n-nubayil</i>	<i>fī-l-mubayil</i>	‘on the mobile’
c. <i>‘āti il-filūs</i>	<i>hāti il-filūs</i>	‘hand me the money’
<i>‘ašīdit šī ‘r</i>	<i>ʔašīdit šī ‘r</i>	‘a poem’
<i>‘add yimsaḥ</i>	<i>ḥadd yimsaḥ</i>	‘someone wipes’

Two noteworthy vocalic features are vowel raising, which targets a non-final short low vowel /a/ followed by another /a/ or /ā/, and produces a high front /i/ (2a), and vowel harmony, which duplicates high /i, u/ to an /a/ in the preceding syllable (2b).

(2) Limbi Arabic	Underlying	Gloss
a. <i>šāyif il-wigāha</i>	<i>šāyif il-wagāha</i>	‘see the elegance?’
<i>wilād ḥilāl</i>	<i>wilād ḥalāl</i>	‘decent people’
<i>xīmas daʔāyiʔ</i>	<i>xāmas daʔāyiʔ</i>	‘five minutes’
b. <i>‘a š-šīrīḥa</i>	<i>‘a š-šarīḥa</i>	‘on the chip’
<i>‘āyiz tiʔgīl</i>	<i>‘āyiz taʔgīl</i>	‘I want an adjournment’
<i>nufūx tāni</i>	<i>naḥūx tāni</i>	‘another crown of the head’

These patternings can be perceived partly as an over-adaptation of correspondence rules which connect Educated Cairene Arabic with higher or lower registers of Egyptian Arabic, or as genuine manifestations of the “street vernacular” spoken by the socially inferior classes in inner-city slums. My analysis shows that il-Limbi’s phonological system is grounded in both, and that it is the interplay between various registers that creates humor in the films. In this way, the social significance of il-Limbi is established primarily through language. The analysis also provides insights into the social characterization of speech in Egyptian pop culture productions, starting with phonology – a characterization which echoes and represents and caricatures actual representations of language in contemporary Egypt.

Length and lexical stress in Palestinian Arabic /a/

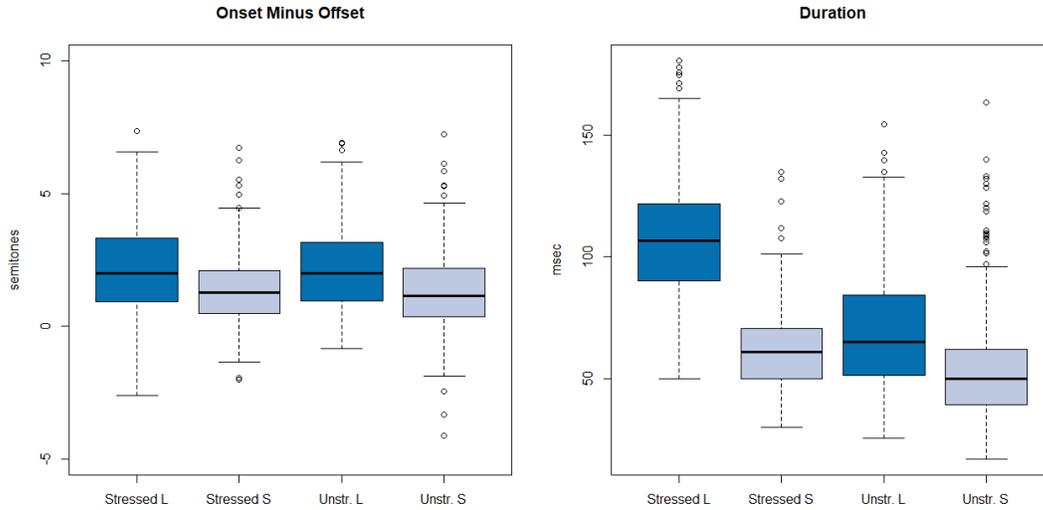
Niamh Kelly (American University of Beirut)

Palestinian Arabic is a variety of Levantine Arabic, which also includes Syrian, Lebanese and Jordanian. Phonetic research on prosody in Arabic includes work on lexical stress and focus in Jordanian Arabic (de Jong & Zawaydeh 1999, 2002) and on the intonation of focus in Egyptian Arabic (Hellmuth 2006a,b) and Lebanese Arabic (Chahal 2001, Chahal & Hellmuth 2014). No phonetic research appears to have yet been conducted on the acoustic correlates of stressed versus unstressed vowels in Palestinian Arabic (but Hall (2020) is examining the correlates of length). Cross-linguistically, lexical stress is often represented by longer duration, higher intensity and spectral tilt, and more peripheral quality (Gay 1978, Beckman 1986, Sluijter & van Heuven 1996, Gordon & Roettger 2017).

The current study examines the acoustics of lexical stress in long and short /a/ in Palestinian Arabic. (The data were kindly provided by Hall (2017).) Target words were disyllabic or trisyllabic with penultimate stress, with both the stressed and unstressed vowels either long or short /a/. The goal was to determine the correlates of lexical stress, as well as to determine whether length interacts with stress, that is, whether long vowels that do not carry stress have the same acoustic correlates as short unstressed vowels. Example words are ['ʒa:bha, 'ʒabha, sa'baqna, sa:'baqna]. In total, we examined 5 stressed and 3 unstressed long vowels, and 6 stressed and 6 unstressed short vowels.

Recordings of 14 speakers (7F, 7M) producing 14 target words were analysed. Each word was produced twice in sentence-medial position. Praat scripts measured mean f0 (semitones), f0 range (st), duration (msec), mean intensity (dB), F1 & F2 (Hz), spectral tilt (H1-H2) and onset minus offset f0 (st) to examine whether the f0 contour of the vowel was rising or falling. The total number of tokens was 1075. These were subjected to a linear mixed effects regression analysis (using the *lmer* function in R (R Core Team, 2019)). The best model for each measure was chosen by model comparison using the *anova* function. The reference levels were the stressed vowel and the long vowel. Speaker and token were random factors.

The best models for each measure (Table 1) showed that the acoustic correlates of lexical stress are higher mean f0 and intensity, longer duration, lower spectral tilt (more modal voice quality), lower F1 and lower F2 (a higher and backer vowel). Long vowels were correlated with longer duration, wider f0 range, higher mean intensity, higher F2 (fronter vowel), lower spectral tilt, and a more falling contour. This suggests that long vowels have effects that occur regardless of stress, revealing that unstressed long and short vowels are not the same. Further work is examining whether the correlates of stress and length are affected by a following pharyngealised (“emphatic”) consonant. This work adds to phonetic research on lexical stress in Arabic, in particular an understudied variety.



Measure		β	SE	t	p
Mean f0	Intercept	89.2	1.04	86.2	<0.001*
	Stress	-1.35	0.1	-13.6	<0.001*
F0 range	Intercept	2.56	0.18	14.15	<0.001*
	Vowel Length	-0.63	0.15	-4.15	<0.001*
Onset - Offset	Intercept	2.22	0.28	7.94	<0.001*
	Vowel Length	-0.8	0.22	-3.71	<0.001*
Mean intensity	Intercept	77.9	1.4	56.38	<0.001*
	Stress	-1.8	0.19	-9.2	<0.001*
	Vowel Length	-0.9	0.24	-3.6	<0.001*
F1	Intercept	544.8	29.5	18.5	<0.001*
	Stress	30.1	7.1	4.2	<0.001*
F2	Intercept	1585	57.2	20.1	<0.001*
	Vowel Length	-75.9	16	-4.7	<0.001*
	Stress	78.4	10.4	7.5	<0.001*
Spectral tilt	Intercept	2.12	0.92	2.3	0.03
	Vowel Length	2.27	0.33	6.99	<0.001*
	Stress	2.65	0.24	11.1	<0.01*
Duration	Intercept	106.1	4.1	25.6	<0.001*
	Vowel Length	-45.29	3.1	-14.6	<0.001*
	Stress	-33.03	3.68	-8.9	<0.001*
	Interaction	23.53	4.7	5.04	<0.001*

Table 1. Linear mixed effects regression results for the best model for each measure.

Cross-dialectal variation in phonetic realisation of the emphatic contrast in Arabic fricatives

Mahmoud Alsabhi (University of York), George Bailey (University of York) & Sam Hellmuth (University of York)

Previous research has revealed a typological divide among Arabic dialects in the phonetic realisation of the phonological contrast between plain and ‘emphatic’ (i.e. pharyngealized) plosives (such as /t~/t^ʕ/): some dialects show an acoustic difference in the plosive (in VOT) and in the following vowel, but other dialects show a difference only in the following vowel (Bellem, 2014).

This study explores whether there is a parallel divide among dialects in the phonetic realisation of the same phonological contrast in fricatives, i.e. between /s~/s^ʕ/. Read speech production data were collected from up to 12 speakers (6M/6F) in each of eight Arabic dialects as part of the Intonational Variation in Arabic (IVAr) project (Hellmuth & Almbark, 2017). Stimuli comprised 24 lexical items in total: 2 target consonants (/s~/s^ʕ/) x 2 word-positions (initial/medial) x 6 following vowel contexts (/i a u i: a: u:/). We report here analysis of a pilot dataset (N=121) of six lexical items with a word-initial target consonant, from 22 male speakers across four dialects: Egyptian (*egca*) and Syrian (*syda*) [dialects with no /t~/t^ʕ/ VOT distinction] versus Jordanian (*joka*) and Kuwaiti (*kwur*) [dialects which have a /t~/t^ʕ/ VOT distinction]. Segment labels were force-aligned to the acoustic signal (Gorman, Howell, & Wagner, 2011) then manually corrected by the first author. A Praat script was used to extract acoustic measurements: duration of the fricative, centre of gravity (COG), intensity and spectral peak location (Hz) at the mid-point of the fricative, as well as the first three formants (F1, F2, and F3) in following vowels. The results were modelled using linear mixed effects models for each acoustic measure in turn as dependent variable with a parallel model structure: DV ~ condition * dialect + vowel + (1 |speaker).

Pilot results (Fig. 1) suggest few differences between the dialects in the phonetic correlates of the fricatives themselves; *egca* and *joka* show somewhat higher values of COG and spectral peak location than are observed in *kwur* and *syda*. The only clear differential marking of the contrast across dialects is observed in the following vowel, and shows a pattern which does not match the typological divide found for plosives. Instead, *egca* and *joka* form a group in showing greater F2 lowering than the other two dialects, and *kwur* and *syda* form a group in showing greater effects on F3; *egca* is alone in showing no use of F1 to mark the contrast. In the full paper we will present results from an enlarged corpus subset, with comparison to emerging results for plosives from the same speakers and interpretation in the light of the overall system of each dialect.

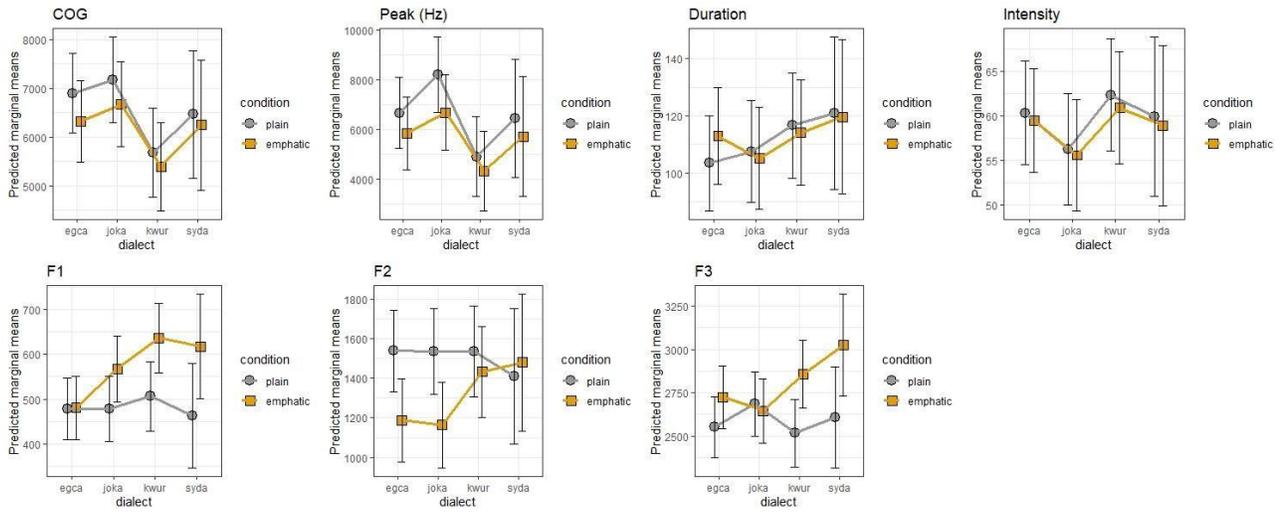


Fig. 1: Predicted marginal means of linear mixed effects models for acoustic measurements taken in the fricative (top row) and in the following vowel (bottom row) by dialect and by condition (plain/emphatic).

Laryngeal cues and the phonological emphatic contrast in spoken Arabic dialects

Latifa Alkuwaiz (University of York), Rana Almbark (University of York) & Sam Hellmuth (University of York)

Arabic dialects have been found to display differences in the mapping of voice onset time (VOT) as a cue to the plain vs. emphatic phonological contrast in Arabic plosives (Bellem, 2007), with some dialects having a three-way distinction (voiced – voiceless emphatic – voiceless plain) and others a two-way (voiced – voiceless). This paper aims to confirm this variation and clarify its nature through examining VOT (and F1/F2 in following vowels) in production data from eight Arabic dialects. Four plosives are examined: plain /t/ and /d/ and emphatic /t̤/ and /d̤/.¹ This is the first cross-dialectal examination of parallel data, to our knowledge, that directly compares VOT patterns for more than two dialects, including previously unexplored dialects like Tunisian and Omani. The results point to a continuum of variation, which supports interpretation of this cross-dialectal variation as evidence of a sound change in progress (Bellem, 2014).

Data from 88 speakers in total across eight Arabic dialects (Moroccan, Kuwaiti, Jordanian, Iraqi, Omani, Tunisian, Egyptian, Syrian) were used for this experiment. Recordings were made on location in the city of origin from 12 speakers in each dialect (6F/6M), except for Iraqi (6F/4M) and Syrian (3F/3M) who were recorded in Jordan. The stimuli comprise real/nonsense words to create emphatic/plain minimal pairs (e.g. /ta:r/~/t̤a:r/, /da:r/~/d̤a:r/) followed by long/short vowels (/a i u a: i: u:/) in different positions (word-initial/-medial), presented in Arabic-script in a dialect-specific carrier phrase. The location of the burst and voicing onset were labelled manually, and VOT/F1/F2 extracted using a Praat script, yielding 3268 fluent tokens to analyse. Data are explored through visualization of VOT values of all tokens to evaluate these differences (see figures below) and a series of linear mixed-effects models (LMM) in R.

The pattern of cross-dialectal variation in VOT is found in the extent of voicing lag in /t~/t̤/. Visualization of the raw data (*figure.1*) suggests a continuum of change, especially for potentially ‘in-transition’ dialects in the middle (Tunisian/Omani) where gender differences are also seen (with male speakers leading the change towards collapse of the VOT distinction). For the voiceless subset, LMM results (*figure.2*) confirm this pattern of variation and change in VOT with some dialects showing a complete overlap of VOT values for plain/emphatic voiceless plosives. Despite this overlap, the extent of coarticulatory effects of emphasis in the adjacent vowels remains unaffected: F2 lowering and F1 raising is still present in the emphatic vowels across all dialects, supporting claims of the primacy of the vocalic cue to the plain/emphatic contrast (Al-Masri & Jongman, 2004; Jongman, Herd, Al-Masri, Sereno, & Combest, 2011; Khattab, Al-Tamimi, & Heselwood, 2006).

¹ A subscript dot rather than IPA diacritics is used to distinguish emphatics from plain counterparts. This is to avoid assumptions about phonetic realization. For the same reason, ‘emphatic’ rather than e.g. ‘pharyngealized’ is used.

Figures

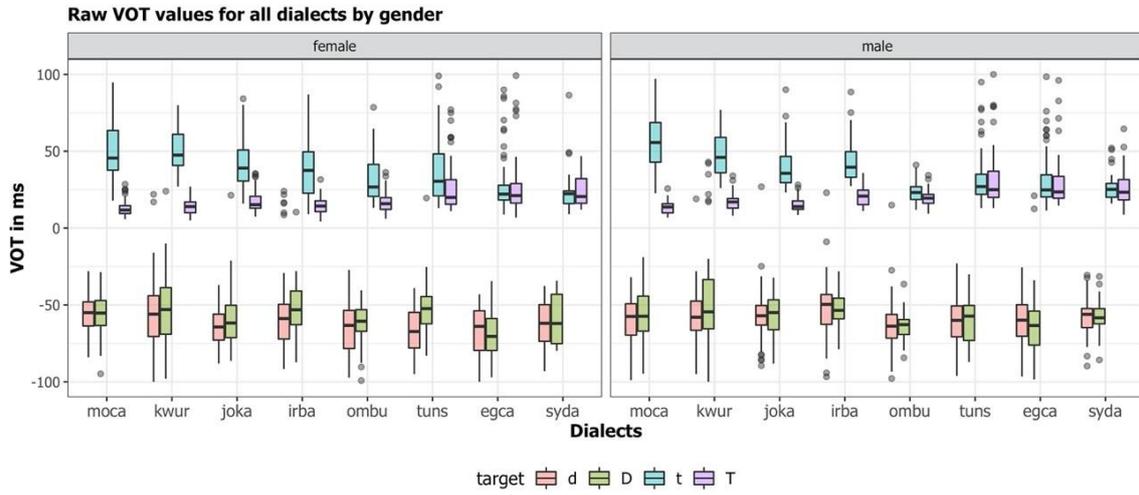


Figure.1

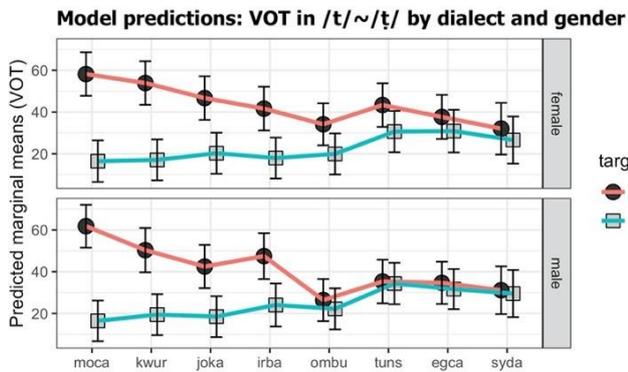


Figure.2

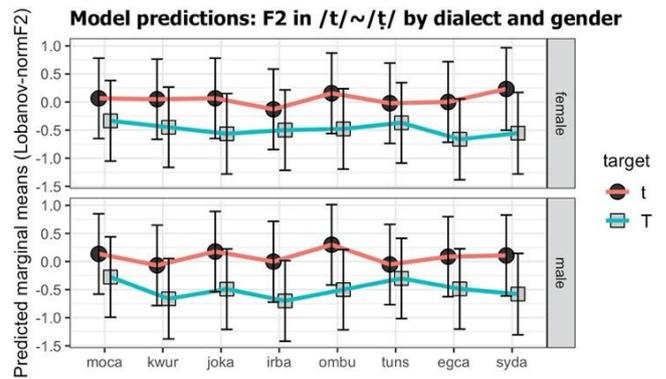


Figure.3

The French voiced uvular fricative in Hedi Zaiem's speech: a pilot study

Lamia Trifi (The University of Texas at Austin)

French is considered a second language in Tunisia and has a crucial role in everyday speech (Miled, 2007). Distinctive phonological features can be perceived in the discourse of men and women when using French loanwords. Men tend to roll the French voiced uvular fricative, i.e. produce an alveolar trill (Lahrouchi, 2018; Maume, 1973) to avoid sounding effeminate and unpatriotic (Stevens, 1980). Women, however, use the correct production of [ʁ] to sound educated and prestigious (Mejri, 2012; Walters, 2011; Dhaoudi, 2009; Marzouki, 2007). Stevens (1980) argues that women try to master their pronunciation of French and the way of achieving this goal is to produce [ʁ] when using French words in their speech which is not well received by men. Yet, we can perceive that men do produce [ʁ] in their language. Hence, this present work is addressing this issue by investigating a Tunisian TV presenter's use of the voiced uvular fricative in his speech. Hedi Zaiem's stylistic variation is analysed in an attempt to capture instances of speech accommodation (Giles, 1973). The focus was set on gender and its influence on Zaiem's production of [ʁ].

The data originates from seventeen videos found on Youtube. These videos were interviews of Tunisian celebrities, eight females and seven males. Thirty-one tokens were collected auditorily, where the two rhotics were produced. After implementing the auditory perception, Zaiem's production of [r] and [ʁ] were compared acoustically, inspired by the methodology of Hirson and Sohail (2007) who examined the variability in Punjabi-English bilinguals' rhotics. For a clear distinction between the rhotics, attention was drawn to an F2 lowering combined with a higher F3 and frication (Thomas, 2011), in the case of [ʁ], and an F2 lowering with a presence of trilling cycles visible on the spectrogram (Jauriberry et al., 2015). Consequently, the auditory and acoustic perception helped determine whether Zaiem converged or diverged from his guests. This pilot study of the nature of rhotics indicated two things. First, Zaiem and his male guests (contradicting previous research) tend to use the alveolar tap instead of the alveolar trill. Second, they switch between the correct production of the fricative and the alveolar tap when speaking Tunisian Arabic embedding French loanwords, i.e., they use both productions randomly. Zaiem produced ten out of twelve tokens with [r] and two with [ʁ] in the presence of his male guests. He also produced twelve tokens out of nineteen with [r] and seven with [ʁ] in the presence of his female guests. These results imply that Zaiem accommodated his speech and tended to converge to women (Giles, 1973). One particular finding was challenging to analyse: Zaiem regularly used Modern Standard Arabic while interviewing his guests. We can interpret this use as a way of diverging from his guests to stay truthful to his journalistic personae.

Acoustic Analysis of Blocking Consonants and Emphasis Spread

Laura Rose Faircloth (The University of Texas at Austin)

BACKGROUND In the phonological process of *emphasis spread* in Arabic, emphatic consonants affect adjacent and non-adjacent vowels within the same word [10]. In Northern Palestinian Arabic, it is claimed that rightward emphasis spread only affects the following low vowel and that [+high] blocking consonants (\int ʒ j w ɪ u) further restrict rightward spread [3]. Central /a:/ backs to low /ɑ:/ in Example (1), while /a:/ is not backed in Example (2) because of /w/. This backing manifests acoustically as F2 lowering in vowels neighboring emphatic consonants [1, 2, 4, 5, 11].

(1) /t^ʕa:za/ → [t^ʕɑ:ʕza] *fresh* [3 p. 485]

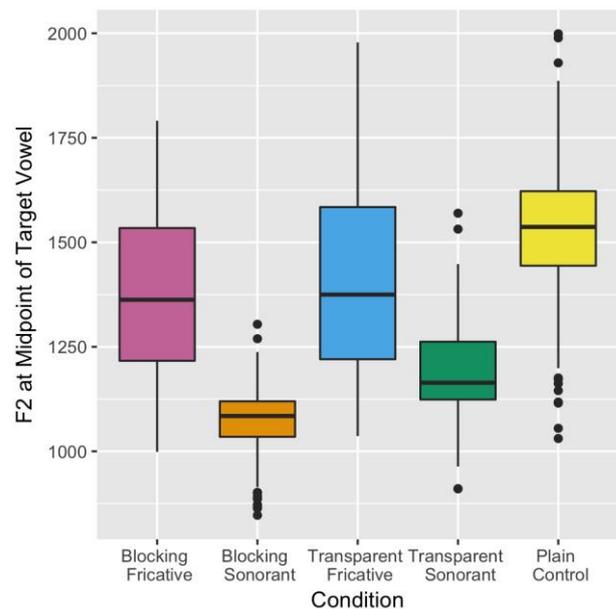
(2) /t^ʕwa:l/ → [t^ʕwa:l] *long* [3 p. 485]

Acoustic work found that the rightward effects of emphatic consonants can spread beyond the following vowel, based on F2 lowering in non-adjacent vowels [4, 11] The present study examined if the blocking consonants restricted rightward long-distance effects of emphatic consonants, as well as the role of the manner of the blocking consonant. F2 is expected to be lower in words with emphatic consonants, but F2 is expected to be higher if the vowel is separated from the emphatic by a blocking consonant.

METHODS Eight native speakers of Nazarene Arabic recorded meaningful words of the pattern C^s₁V₂C₂V_sC₃, with a word-initial emphatic consonant (C^s₁) with a target vowel in the next syllable (V₂). The intervening consonant (C₂) was either blocking \int w / or transparent l ɣ /. To examine the role of manner, one of each was a sonorant w l / and the other a fricative \int ɣ /. These were compared to control items, which did not have an emphatic consonant. F2 was measured at the midpoint of each target vowel (V₂).

RESULTS There was a significant effect of CONDITION (χ^2 (4) = 132.4, $p < 0.0001$). All vowels in the EMPHATIC conditions exhibited F2 lowering in comparison to the CONTROL, though this lowering was smallest with the fricative blocking consonants. F2 lowering was greater in vowels separated from the emphatic consonant by a sonorant consonant than those separated by a fricative. The blocking consonants \int w / did not block the long-distance effects of emphatic consonants on F2.

CONCLUSIONS This study indicates that the emphatic consonants affect non-adjacent vowels within the same word and that the blocking consonants reported in Davis [3] did not limit these effects. The effects of emphatic consonants were smaller with the fricatives \int ɣ / than with sonorant l w /. This result may be due to differences in the dialect studied from Davis [3], who found that w / was a blocking consonant in Northern, but not Southern, Palestinian Arabic. However, fricatives have been found to be more resistant to co-articulation cross-linguistically [8, 9]. The greater effect of emphatic consonants with sonorants suggests that emphasis spread is more restricted by [-sonorant] than by [+high] [3, 6, 7].



Depalatalisation of /gj/ and /kj/ in the Ḥarbi dialect in Medina: patterns of variation and change

Mohammad Alrohili (Tabuk University)

This research investigates sociolinguistic variation and change in the dialect of the Ḥarb tribe in Medina (west of Saudi Arabia). The dialect belongs to the north-west family of dialects. Two of the salient features of this dialect are examined in this research: 1) the traditional palatalised [gʲ] and the innovative velar stop [g] as two variants of the variable (g) and 2) the traditional palatalised [kʲ] and the innovative velar stop [k] as two variants of the variable (k). These two features are investigated in this research regarding three social factors: age (young, middle-aged and old), gender (male, female) and levels of contact (high, low). To date, these features have not been investigated as sociolinguistic variables in the city of Medina. The analysis presented in this study comes from large-scale research, and the data were obtained through sociolinguistic interviews with 43 native speakers distributed over three age groups from both genders. The level of speakers' social contact (low/high) was also measured. Regarding linguistic factors, the data were coded for the preceding and following sounds. Data were analysed using Rbrul statistical software. Concerning the variable (g), women were found to be ahead of men in using the innovative [g] at 82% and 67%, respectively. Contact was returned as the most significant factor, with high-contact speakers leading low-contact speakers in using non-traditional variants at 96% and 53% for low-contact speakers. Age was found to be significant, with the young and middle-aged being the most innovative groups at 80% and 74%, respectively. The oldest group favoured the palatalised [gʲ] at 66%. Respecting linguistic factors, Rbrul returned the preceding and following environments as significant. The variable (k) at the beginning was treated as two variables, as it occurs in the stem and the suffix. Subsequently, stem and suffix tokens were grouped and analysed together due to the small number of tokens in the suffix. The results show that women (90%) and speakers with a high level of contact (96%) led the change in using the innovative feature [k]. Regarding age, there was no great difference in the use of the innovative variant [k] across all of the age groups. Middle-aged and young speakers scored 87% and 85% respectively, whereas old speakers scored 80%. The results show the traditional palatalisation variants [gʲ] and [kʲ] are changing in the speech of Ḥarbi clans in Medina and there is a tendency to use the innovative and supra-local forms [g] and [k].

Emphasis and the Temporal Organisation of Consonant Sequences in Arabic

Khalid Alsubaie (University of Leeds)

The main aim of this study is to acoustically examine the effect of emphasis on the temporal organisation of consonant sequences. It has been reported that order of place of articulation influences the temporal organisation of consonant sequences, particularly in the case of stops (e.g., Byrd, 1994 for English, Alsubaie, 2014 for Najdi Arabic, Ghummed, 2015 for Libyan Arabic). Anterior-posterior constrictions like /tk/ tend to be shorter in duration than posterior-anterior constrictions in a cluster like /kt/. Emphatic consonants, which are produced with a secondary constriction in the posterior vocal tract, are considered. This study examines whether the secondary articulation in these sounds affects the temporal coordination in consonant sequences that involve such sounds in Arabic.

Najdi Arabic has been chosen to be the examined variety because emphatic consonants (t^s , δ^s , s^s) occur in this dialect and it permits consonant sequences involving these consonants word-initially, word-finally, and across a word boundary. Sixteen native speakers of Najdi Arabic produced a list of words in (C)C#C(C) sequences that include at least one of the emphatic coronals / t^s /, / s^s / or / δ^s /. Recorded data were acoustically analysed using Praat speech analysis software. Various acoustic correlates were examined including hold phase duration for the plosive, frication duration for the fricatives and duration of the interconsonantal interval (ICI). Measurement results were processed statistically with speaker identity as a random factor and the context (whether plain or emphatic) as the crucial fixed factor.

Preliminary results suggest that there is a significant effect of emphasis on the temporal organisation of consonant sequences in Najdi Arabic. The manner of articulation and order of place of articulation play a role in this effect.

Dorsal Rhotics in Arabic: Evidence From qəltu-Arabic

Uthman Aldahook (Durham University)

Traditional scholars of Arabic describe 8th-century Arabic rhotics as apical resonates (AlKhalīl ed. al-Maḥzūmī & al-Sāmmarrā'ī 1988; Sībawayh ed. Hārūn 1982:448; Ibn Jinnī ed. AlSaqqa 1954:1/60- 61). Recent articulatory accounts on Arabic rhotics provide more details on the manner in which rhotics can be produced across the modern varieties of Arabic (Nasr 1966; Al-Ani 1970; Shaheen 1979; Anani 1985; Khattab 2002; Watson 2002; Bellem 2007; Sylak-Glassman 2014). These accounts demonstrate additionally that apical rhotics in Arabic can be plain /r/ or pharyngealized /rʕ/.

Evidence for the existence of dorsal rhotics [ɣ~ʕ] in Arabic dates back to Abbasid times (Al-Jāhīz ed. Hārūn 1986; Jastrow 2006). An overwhelming part of the Arabic grammar study tradition neglects and stigmatizes dorsal constrictions of Arabic rhotics (Al-Kindī ed. Al-Tayān 1985; Al-Jāhīz ed. Hārūn 1986). This prescriptive approach marginalizes speech patterns potentially peculiar to substratum areal features. After perhaps the 10th century, Arabic started to acquire new native speakers in diaspora (Blanc 1964: 167, 202). Some of those native speakers started writing in the colloquial local forms which later served contemporary linguists with features of spoken Arabic in designated historical periods of time (Versteegh 2014: 170; Levin 1975). More importantly, such pieces of evidence help to provide information about dorsal rhotics.

A synchronic dorsal-apical-r merger in some varieties of Arabic suggest that Arabic rhotics are produced in two points of articulation: apical and dorsal. Dorsal rhotics are part of a bundle of hallmark features for the Tigris subgroup of qəltu-Arabic of the phylum Mesopotamian Arabic (Jastrow 2006). As far as Arabic-speaking communities at large are concerned, dorsal rhotics are also reported in Maghrebi varieties of Arabic that cluster together with areal linguistic features: (Morocco) in Fez; and Meknes and marginally in Tetouan (Stroomer 2004). A dorsal r-sound was also attested in Eastern Kabylia (Algeria), strictly in the town of Djidjelli (Marçais 1956).

The aim of this paper is to demonstrate further that the dorsal-apical-r merger exists in Arabic from empirical evidence. This study analyses r-sounds from sound recordings of Jewish Iraqi Arabic speakers in London, Canada and Israel¹. The preliminary findings suggest a demonstrated intra-speaker and inter-speaker variation in apical-dorsal-r. For instance, [ˈhəlɪkəʔ.tər] ‘helicopter’ and [ˌtʕɑj.jɑːʕˈvɑt] ‘planes’; [ɛsʕsʕuˈʕɑː] ‘sketch’ and [rʕsɑˈmu] ‘(he) sketched (it)’. Thus, the study also contributes to the need for a dorsal-r classification and broader [ɣ~ʕ] areal typology for Arabic rhotics.

¹ From the corpus of spoken data recorded by Eli Timan, documentation of the Jewish Iraqi spoken language, 2006 - 2007 [computer file]. London: Endangered Languages Archive (ELAR) [distributor], November 2010. <http://elar.soas.ac.uk/deposit/timan2008jewishiraqi>

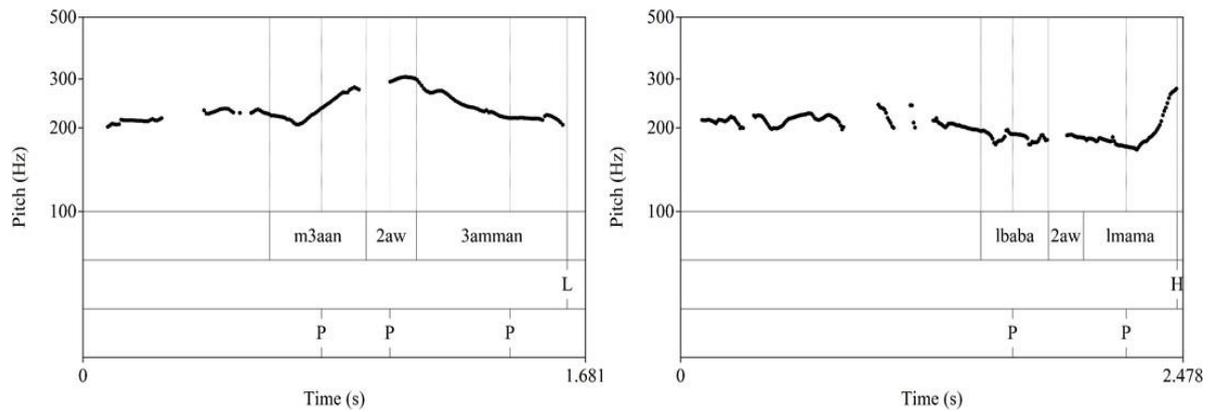


Figure 1. Pitch traces of nuclear contours (X or Y) of altqs (left) and dynqs (right)

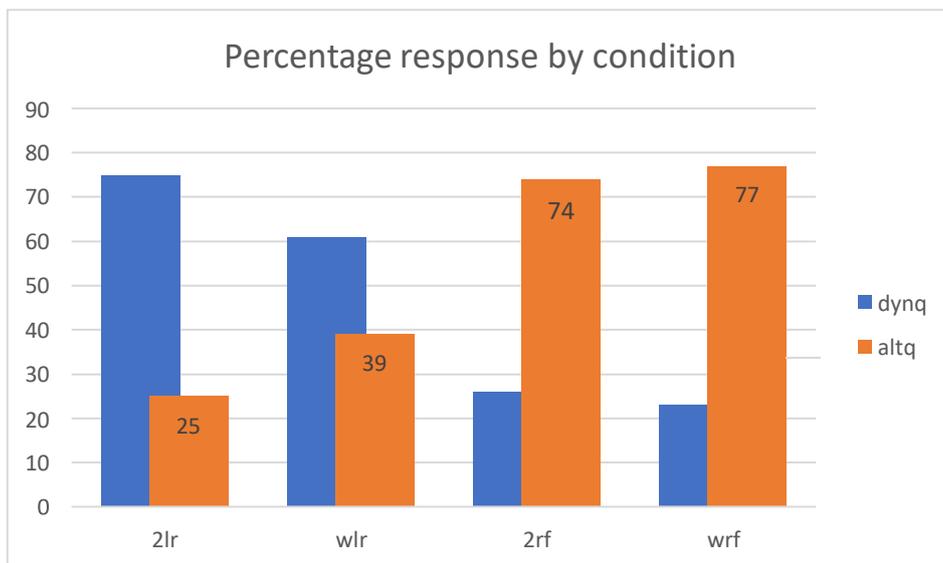


Figure 2. Percentage of responses in the four conditions in the perception experiment

On the Relationship between Morphology and Reading Accuracy among Children in Israel

Duaa Abu Elhija Mahajna (Bar Ilan University) & Abeer Asli-Badarneh (Bar Ilan University)

Arabic-speaking children growing up in Israel face a daunting task in learning to read in two languages, which is made more difficult by Arabic diglossia, whereby the morphology of the Spoken Arabic (SA) differs from that of Modern Standard Arabic (MSA). The present study investigates reading accuracy and speed performances of first grade students in northern Israel divided into four experimental groups: Arabic monolinguals, Hebrew monolinguals, Arabic- Hebrew sequential bilinguals (Arabs), and Hebrew-Arabic sequential bilinguals (Jewish). The aim of the study is to measure and compare fluency in reading performance among these four groups in the native and second language and to examine how morphological performance influences reading ability in the two languages. The study also aims to assess the differences in spoken and standard language performances in Arabic among the experimental groups. The study involved 200 participants, monolingual and bilingual 1st grade students aged between 6-7. The participants were recruited from four schools in the North of Israel, two monolingual schools (one Hebrew, one Arabic) and two bilingual schools.

To see the relationship between reading ability and morphological performance several measures and tasks were conducted. Morphological abilities were assessed by using five tasks on inflectional and derivational morphology. Part of the morphological tasks for Arabic included MSA words and another part – SA words. New combined variables of derivational and inflectional morphology were re-created by combining all morphological tasks in L1 and L2. Texts for reading assessment (both Arabic and Hebrew versions) were taken from the first-grade curriculum in the respective schools and aimed to examine the child's reading level. The texts were unknown to the children but contained no unlearned letters. Each text contained 40 words (in MSA or Hebrew, respectively). MSA and SA proficiency was assessed by a vocabulary knowledge test and spoken word identification, whereby subjects' familiarity with the target word was determined.

The results were analyzed using a one-way ANOVA test. The main results show that, with respect to the native language, Hebrew-speaking monolinguals read accurately and faster than both Arabic-speaking monolinguals and Hebrew-speaking bilinguals ($p < .001$). Arabic-speaking monolinguals read worse and slower than Arabic-speaking bilinguals, Hebrew-speaking monolinguals and Hebrew-speaking bilinguals ($p < .001$).

As for reading in the second language, Hebrew-speaking children revealed better fluent reading in their native language compared to their second language, while Arabic-speaking children showed a more balanced reading performance in the two languages.

Post hoc follow-up comparisons (Scheffe tests) of MSA performance in the three pertinent experimental groups indicated that Arabic-speaking monolinguals performed similar to Arabic-speaking bilinguals and both performed better than Hebrew-speaking bilinguals in MSA ($p < .001$). In SA performance, Arabic-speaking monolinguals show similar performance to Arabic-speaking bilinguals, while both performed significantly higher than Hebrew-speaking bilinguals. Finally, higher ability in derivational and inflectional morphology correlated with reading accuracy in the native language, while only derivational morphology correlated with reading time in the native language. In the second language, reading was correlated with higher ability in derivational morphology both in accuracy and time measures.

Competing patterns of possession: Evidence from loanwords in Palestinian Arabic

Lior Laks (Bar Ilan University)

This study examines variation in marking possession on loanwords Palestinian Arabic (PA). Such variation is demonstrated below.

- (1) istaxdem **il-kami:ra btaʕtak** aw **tilifo:n-ak** 'use
your camera or your phone'

<https://sandbox.altv.com/blogs/videotipsarabic/>

While *tilifo:n* 'telephone' takes the suffix *-ak*, *kami:ra* 'camera' takes the genitive exponent *btaʕtak*. The study examines variation between morphological and periphrastic possession (also known as synthetic vs. analytic). Many colloquial dialects show variation between the two patterns, e.g. *kta:b-ak* vs. *il-kta:b btaʕtak/tabtaʕak/ta:ʕak/ʕe:tak*¹ 'your (masc. sg.) book' (see Eksell Haring 1980, Pavla 1991, Borg 1994, Talmon 2002, Owens 2002, Teie- Cheikh 2010, Sayahi 2015, among others). This is more typical in loanwords. I will show that the selection of either strategy can be partially predicated by morpho-phonological criteria. I argue that the more 'foreign' loanwords look like, the less likely they are to be integrated into the morphological inflectional system. I discuss here two such constraints.

Number of syllables. Words that exceed two syllables are more likely to take the periphrastic pattern (*il-instagram btaʕtak* 'your Instagram'). Most Arabic nouns are monosyllabic or disyllabic, unless they consist of affixes. Monosyllabic or disyllabic loanwords can be partially perceived as native, and therefore are integrated more easily and tend to take morphological possession (*imel-ak* 'your email'). Words with more syllables are treated as foreign and the more syllables there are, the more likely they are to take the periphrastic pattern. This is more dominant in borrowed compounds, e.g. 'play station', 'bodyguard'.

Vowel-ending words. Words that end with vowels tend to take the periphrastic possession. This is primarily consistent within words that end with vowels other than *a*, which are relatively rare in PA, e.g. *shambu* 'shampoo', *selfi* 'selfie', *biano* 'piano'. Similarly, the small set of native words that end with these vowels also tend to take periphrastic possession, though to a smaller extent (*masʕa:ri-k* ~ *il-masʕa:ri btaʕtak* 'your money'). In addition, attaching possession suffixes to such loanwords would require phonological alternations. For example, if *-ak* is added to *biano*, either *a* would be deleted or *h* would be inserted (**biano-k/*biano-hak*) in order to avoid the diphthong *oa* (**biano-ak*). Either way, this renders morphological complexity and speakers tend to select the periphrastic pattern (*il- biano btaʕtak*). It is important to note that such criteria represent tendencies, and there are counter examples. The issue here is the differences between types of loanwords and the type of possession they tend to take. In addition, diachronic information also plays a role here. 'Older' loanwords (*telephone*) are more likely to take morphological possession in than 'newer' words (*facebook*). The study adds to previous studies that examine the competition between morphological and periphrastic formation, and offers criteria that play a role in selecting either pattern. The study also sheds light on the degree of integration of loanwords. This provides direct access to word formation and shows how different types of criteria are taken into consideration.

¹ For the purposes of this study, I do not distinguish between the different types of possession particles of periphrastic possession, which also change across different dialects.

Initiated usage of case markers in Standard Arabic text production

Lior Laks (Bar Ilan University) & Elinor Saiegh-Haddad (Bar Ilan University)

The study examines the distribution of case markers in Modern Standard Arabic (MSA) spoken and written texts across different ages. Speakers demonstrate variation in the usage of case markers, as demonstrated in the following spoken text of a fourth grader. The same noun is used in the same syntactic position with and without the case marker *-i*.

(1) rajaʕtu **ʔila l-bayt**...wa-ʕindama rajaʕtu **ʔila l-bayt-i**(Majd) 'I returned to the house...and when I returned to the house'

Case markers are one of the most prominent differences between MSA and colloquial Arabic. Knowledge of case marking is not acquired but learned in school. Moreover, case markers are not commonly used in both written and spoken MSA.

The study examined whether native speakers use case markers. Methodology is based on a cross-linguistic project on developing literacy (Berman&Verhoeven 2002). Participants produced MSA spoken and written narrative texts. Participants were 128 speakers of Palestinian Arabic (PA) of 4th, 7th, 9th and 11th grade. We examine the use of case markers with respect to three dimensions.

Language Development. 4th graders used case markers to a greater extent. This could be explained by the fact that they learn about case markers around this age.

Modality. Case markers were more common in spoken texts. This is surprising, as MSA is typically more written than spoken, and case markers are typically learned in the context of written language. We assume that case markers were used more in spoken texts because participants tried to distinguish between spoken MSA and spoken PA and mark MSA as formal.

Morpho-syntactic factors. Focus is on the distribution of case markers in 4th graders texts, where the usage of case markers was the highest. Distribution can be partially predicted based on systematic guidelines.

- **Lexical category.** Case markers are more common on nouns than adjectives. This suggests that case is perceived as more typical to nouns.

- **Definiteness.** Indefinite nouns demonstrate a higher ratio of case marking. As shown in (2), the speaker used the indefinite noun *film* with a case marker, and the definite noun without one. This is also related to an orthographic factor, namely the use of the letter *alif* in writing, but as we show, the definiteness distinction exists for other case markers.

(2) aħabbat ʔan nuša:hid **film-an**... kunna nuša:hid **al-film**...(Aseel) '(she) wanted that we watch a film....we were watching the film...'

- **Syntactic position.** Subjects tend to be more marked for case. In addition, adjectives marked for case are predicative and not attributive.

- **Bound morphemes.** Most nouns with possessive clitics had case markers. This is because case markers are an internal part of words (*bayuna/baytana/baytina* vs. *baytna* 'our house') and they are perceived as a more integral part of words.

The findings correlate with previous studies that examined spoken MSA in formal speeches and interviews of adults (Meiseles1997, Parkinson1994, Magidow2012, Halberg 2016). The study sheds light on this topic from developmental and methodological perspectives, and the results can also have pedagogical implications for MSA instruction.

The Geographical Distribution of the Definite Article in West-Arabian Dialects

Francesco Grande (University of Turin)

Rabin (1951:1) extensively describes “the ancient Arabic dialects” in use in the Western Arabian peninsula in pre-Islamic times. In his description, Rabin (1951:1-3) applies a “comparative and geographical” approach to “the data preserved by Arab philologists”, finally subsuming these dialects under a West-Arabian group, defined as “a non-literary language”, opposed to “Classical Arabic [...] the language of pre-Islamic poetry”.

The definite article is among the West-Arabian dialectal features described by Rabin (1951:34, 50, 205), who focuses on its realization as (a)m in the Ṭayyi’ and Ḥimyarī tribes: e.g., (a)m-sahm ‘the arrow’. This example clarifies that the sample point of Rabin’s dialectological description is societally specific: a West-Arabian tribal unit rather than a locality (Rabin 1951:15-16). This talk offers a case-study in the realization of the definite article based on a larger sample of West-Arabian dialects. The corpus consists of about thirty idols’ names, selected for three semantic reasons:

- they are unambiguously pre-Islamic, because of their referents
- they are unambiguously definite, being proper names
- they are often speaking names, i.e., proper names that synchronically still function as common names

In particular, some idols’ names combine the semantic property of definiteness with that of speaking name, according to an onomastic pattern where the common name co-occurs with a definite article: e.g., (a)l-nasr ‘the eagle/the Eagle’. This pattern provides the empirical basis for a study of the geographical distribution of the West-Arabian definite article along the following lines.

Data collection procedure

The corpus consists of two similar lists of idols, mentioned in two contemporary Andalusī sources. Al-Mukhaṣṣaṣ (XIII, 104-105) by Ibn Sīdah (d. 458/1066) mentions a list of 32 idols, often including lexical variants. Jamharat Ansāb al-‘Arab (II, 491-493) by Ibn Ḥazm (d. 456/1064) mentions a list of 25 idols, with indication of the tribal units worshipping them.

Analysis

- The two lists are compared
- The idols shared by the two lists are singled out
- For each shared idol, any lexical variant of Ibn Sīdah’s list is paired with a given tribal unit of Ibn Ḥazm’s list, whenever the idol worshipped by that tribal unit has a nominal form corresponding to that lexical variant
- In each pair ‘lexical variant-tribal unit’ a realization of the definite article is looked for
- All realizations of the definite article are listed, along with their corresponding tribal units
- From the resulting set of pairs ‘definite article-tribal unit’, those pairs are selected, which feature a West-Arabian tribal unit
- The map of West-Arabian dialects drawn by Rabin (1951:14) is considered, with particular attention to the West-Arabian tribal units used by him as sample points
- The realizations of the definite article and their corresponding West-Arabian tribal units, both resulting from the foregoing comparison of Ibn Sīdah’s and Ibn Ḥazm’s lists of idols, are positioned on Rabin’s dialect map.

Expected results

The geographical distribution of the different realizations of the West-Arabian definite article based on the analyzed sources: e.g., (a)m, (a)l.

The morphological complexity in Arabic spelling

Iyad Issa (Community Development Authority in Dubai)

Spelling poses a challenge to Arabic-speaking learners due to the complexity of the morphological and orthographic system in Arabic. Arabic morphology has been argued to play a critical role in spelling since its morphological operations are built on a system consisting of a root that is interlocking into different patterns of vowels to form different categories of words. In addition, Arabic orthography is considered to be loyal to the morphographic principle (Ravid, 2012), where morphemes correspond to graphic representation regardless of the pronunciation, especially in the non-vowelized texts. Knowledge of predictable morphological forms depending on the morpho-orthographic analogy can play a role in improving the spelling processes and spelling of new words (e.g., Moats, 1983; Berninger et al., 2010; Deacon and Kirby, 2004; Ravid and Schiff, 2006) and can play a role in facilitating the spelling processes generally.

This study made a detailed classification of spelling errors in a word dictation task, based on morphological structures, undertaken by 107 Typically-developing learners (TD) and learners with learning disabilities (LD) attending the same schools in Jordan. All participants ranged in age from 7 years, 3 months to 15 years, 2 months (grades 2 to 8). The spelling task was made up of 400 common words representing all morphological forms in different conjugations and grammatical classes. The results indicated that TD and LD learners follow a similar pattern of complexity even though the LD group produced more errors than the TD group. Both groups encountered more difficulties in passive voice forms followed by active voice forms while, on the other hand, both groups spelled nouns, verbal nouns and derivations more accurately than verbal forms (active and passive voice). These results extend our knowledge of the complexity that learners tend to encounter in mastering morphological forms which has a significant impact on spelling performance. The results provide additional evidence for the nonlinear growth of morphological knowledge in spelling.

“P” variation in the different perfect constructions across Arabic

Maris Camilleri (University of Essex)

The perfect is primarily understood as a specific construction whose core structure involves an auxiliary along with a participial form (Portner, 2011), which in Germanic/Romance associates with distinct semantic interpretations depending on the presence/absence of adjuncts (e.g. Dowty (1979)). An exemplification is provided in (1) below.

- (1) a. I’ve just got here. Recent Perfect
 b. I’ve got here three years ago. Existential Perfect

In the case of Germanic/Romance, as illustrated from English in (1), it is the same core structure, grammaticalised out of the possessive construction (Heine and Kuteva (2006); Drinka (2017)), that associates with, and expresses, the different readings. In the case in (1), it is the use of (an encliticised form of) the have auxiliary. I here intend to show how this is not true of what one finds in Arabic or in Celtic languages, for instances, with which Arabic draws a lot of parallels, as will be also be shown.

Through a comparative array of Arabic dialectal data, I will demonstrate that while a broad semantic split between the Universal vs. the Existential perfect reading suffices, the syntax does not merely match up in an isomorphic manner. Rather, structural differences which go beyond the mere adjunction of adverbs render finer-grained readings, and that they are additionally indicative of distinct grammaticalisation trajectories. While the classic split that has been discussed so far in the literature on the vernaculars, summarised and brought together in Hallman (2016) and XXXX (2019), makes reference to active participles and perfective forms as expressing an Existential reading, as opposed to designate constructions expressing a Universal reading, I here demonstrate that there is much more to be said about the landscape of perfect expressing construction in Arabic. There appear to be (at least) three specific constructions (beyond other structures involving mere adverbs in relation to perfective, or active participle forms) with varied distributions across the different dialects that express distinct Perfect readings. These are unified by virtue of the obligatory presence of prepositions (Ps), at least, formally; hence “Ps”, since it will be argued that these are functionally/categorially not such. A characterisation of the different Perfect readings, as well as the “Ps” employed, are provided below:

1. Universal perfect: “Ps” *la/il* ‘to, for’ (2), *‘and* ‘at’
2. Existential perfect: “Ps” *la/il* ‘to, for’, *bi* ‘with, in, by’, *ma^ʕ* ‘with’ (3)
3. Recent perfect/Short perfect (Johnson, 1981): “P” *ba^ʕd* ‘after’, *k / ki(f)* ‘like’ (4)

- (2) muna **il**-a xams iyyām bi-l-ħabis
 Muna to-3SGF.GEN five day.PL in-DEF-prison

Muna has been in prison for five days. (Damascene) Syrian: (Hallman, 2016, p. 77)

- (3) aní **má^ʕ**-I ma-fhúm inna aš-šulṭán...
 I with-1SG.GEN PASS.PTCP-understand.SGM COMP DEF-king

I have understood that the king ... Andalusian: Corriente (2008, p. 375)

- (4) **kī** uṣal
 like arrive.PFV.3SGM

He’s just arrived. (Tlemcen) Algerian: Mar,cais (1902, p. 192)

It will be hypothesised that while both the Universal and Existential perfect constructions have grammaticalised from the same type of source structure, the Hot news/Short/Recent perfect reading has developed out of a separate construction. The split finds further support in (quasi) parallels from Celtic, where, while the different languages within the family grammaticalise the same three distinct structures from different sources, intriguingly, the Ps 'to' and 'at', and the P 'after' are present in the expression of the Existential, and the Short perfect readings, respectively.

Participial Structures within the Left Periphery: the Case of the Arabic Active Participle

Mohamed Naji (Moulay Smail University)

Although Arabic is a head initial language with all the lexical heads preceding their subject and arguments, there is a striking asymmetry between participial structures and verbal ones. While the verbal structures have a basic word order with the verbal head in the initial position as in (1), the participial structures have a basic word order in which the subject is in the initial position as the structure with the active participle in (2) shows:

1. ya-*chrabu* l-*walad-u* d-*dawa?-a*
 3.sg.m-drink.imperf det-boy-nom det-medicine-acc
 (The boy takes the medicine.)
2. l-*walad-u* *chaarib-un* d-*dawa?-a*
 det-boy-nom dirink.parti-m.sg det-medicine-acc
 (The boy is taking the medicine.)

This asymmetry raises a question about this difference in the word order into the same language and why does the subject appear in the initial position instead of the lexical participle head in (2)?

In this paper, we will argue for the idea that the Arabic participles differ from lexical verbal heads in lacking an independent tense projection, TP. By assuming the split CP hypothesis of Rizzi (1997), (2014), we will propose that participial structures have only a weak tense category which is fused into the head of FinP category in the lexicon.

Various arguments support the evidence for the absence of an independent tense category in the participial structures. One of them is the impossibility of modality particles to appear with the active participle as illustrated in (3b) down:

3. a. *qad* ya-*ktub-u* zayd-un r-*risalat-a*.
 probable 3.sg.masc-write-indi zayd-nom det-letter-acc
 (Zayd probably will write the letter.)
- b. * *zayd-un* *qad* *katibun* r-*risalat-a*
 zayd-nom probable write.parti.masc.sg det-lettre-acc
 (*Zayd probably writing the letter.)

Furthermore, at the opposite of what happens with active participle in other Semitic languages, like *benoni* in Hebrew which can be inverted or incorporated with auxiliary verbs (Cf. Shlonsky (1997) for more details about *benoni* participle), the Arabic active participle can not be inverted or incorporated in auxiliary verbs. Furthermore, participial structures does not allow the “S O parti” order with the interpolation of the direct object argument between the subject and the participle, as is the case with *benoni* in Hebrew, or in the structures with verbal heads in Arabic. We will assume that although the participle can be negated by a negative marker, this negative marker is not a head marker; it differs from verbal negative heads in being unable to identify a Neg projection. Likewise, participles fall to identify Modal heads too.

The word order between the active participle and its external argument supports the proposal that the active participle checks its categorial [+V] feature of tense in the head Fin^o, while the nominative subject checks the [+D] feature in sepc FinP position.

Lost and newfound features: what urban varieties can tell us about the history of agreement in Arabic

Simone Bettega (University of Turin)

In recent years, a growing number of studies have appeared that focused on the topic of agreement in Arabic (Herin and Al-Wer 2013; Prochazka and Gabsi 2016; Ritt-Benmimoun 2016; D'Anna 2017a, 2017b; Bettega 2017, 2018; Bettega and Leitner 2019). All these works provided a wealth of new data, the analysis of which has brought about the opportunity of better assessing the nature of agreement in Arabic, especially from a typological and comparative point of view (Bettega 2019).

Most of the aforementioned studies, however, were focused on dialects that still retain gender distinction in the plural forms of the adjective, verb and pronoun. So-called “non-distinguishing” dialects have been somehow neglected, partly because already well-documented, partly because agreement patterns in these dialects are perceived as “less complex”, and therefore in need of less careful scrutiny.

Remarkable variation, however, exists among these varieties, and, to this day, reliable studies documenting such variation are few (Belnap 1991 on Cairene; Hanitsch 2011 on Damascene; Prochazka and Gabsi 2016 on Tunis Arabic; Brustad 2000 on Cairene, Damascene, Urban Moroccan and Kuwaiti: only Belnap provides a quantitative analysis of his data).

This presentation has two main objectives: the first one is to offer a corpus-based, statistical analysis of agreement patterns in well-documented varieties of Arabic (such as Urban Moroccan or Damascene) whose agreement behavior is still, however, known at a rather superficial level. The second objective is to investigate what these “non-distinguishing” varieties can tell us about the history and evolution of gender and agreement in Arabic.

Two main hypotheses will be put forward, namely:

(I) that the loss of feminine plural agreement in spoken Arabic was not the result of phonetically-conditioned morphological erosion, but was brought about by a shift in the speakers' patterns of use (possibly induced by ecological/environmental factors, such as a general change in their living conditions);

(II) that the loss of feminine plural agreement inevitably caused a rearrangement of the whole gender/agreement system, one that led to different outcomes in different varieties. Some dialects (type A) simply maintained the original variation between feminine singular (non-individuated) and plural (individuated) agreement, replacing the original feminine plural forms with the new genderless plural. Other dialects (type B) disposed of both feminine plural and feminine singular agreement, leaving the originally masculine (now genderless) forms as the only plural agreement option. It is probable that these dialects first passed through a type A stage. A third theoretical option (type C) exists, that is the replacement of all plural agreement options with feminine singular forms. While this is far from being the actual state of affairs in any known dialect, it can be noted that feminine singular agreement with plural controllers appears to be statistically more common in Cairene Arabic than it is in other dialects (Belnap 1991).

Grammatical Development of the North African Purposive

Mike Turner (University of North Carolina Wilmington)

In this study I use synchronic data from modern Moroccan Arabic to sketch a probable pathway for the historical emergence of the North African purposive structure *bāš*. Nearly ubiquitous in North African Arabic, early attestations of dialectal *bāš* with a meaning ‘in order to’ are found in a 15th-century Judeo-Arabic mercantile text (Wagner, 2017, p. 79) and a 14th-century Maghrebi poem known as *Mal‘abat al-Kafīf az-Zarhūnī* (Benchrifa, 1987), which was cited in part by Ibn Khaldoun. Considering it is also present in Maltese, such evidence points to the purposive *bāš* as an important early North African isogloss, but the processes that led to its establishment have not seen explicit study.

I argue here that it is possible to reconstruct the development of *bāš* by looking at the form’s semantic and syntactic variability in living varieties. Taking Moroccan Arabic as a case study, I use a corpus from Post (2015) alongside my own data to establish that *bāš* is not simply a purposive, but rather a polyfunctional morpheme with a number of discrete grammatical roles, among them:

Interrogative:

- (1) **bāš** tā-tt’əlləf hād d-dūla?
 INT IND-be.composed DEM ART-country
 ‘of what is this country composed?’

Relativizer:

- (2) ‘əndu xālfīyāt **bāš** wəlla hākkāk
 have.3MSG experiences REL became.3MSG thus
 ‘he has experiences through which he got to be like that’

Purposive:

- (3) kā-nqra **bāš** nžīb n-nūqāt w-šāfi
 IND-study.1SG PURP bring.1SG ART-grades and-enough
 ‘I study in order to get good grades, that’s it’

Complementizer:

- (4) hīya gālt -līna **bāš** nbəhtu
 3FSG tell.3FSG -1PL.DAT COMP research.1PL
 ‘she told us to do research’

Nominalizer:

- (5) **bāš** ykūn ‘əndək Facebook wə- nta mzūwəž māši hīya...
 NOM be.3MSG have.2SG Facebook and- 2MSG married NEG 3FSG
 ‘having Facebook when you’re married is not...’

I propose that the above variability is not random, but rather is the patterned result of a historical process of grammatical accretion whereby the semantic range of *bāš* was successively extended to new meanings while leaving earlier ones in place. This process was likely driven by reanalysis of the morpheme in ambiguous contexts, corresponding with what Heine (2003) calls “bridging stages” of grammaticalization. Ambiguous instances of *bāš* abound in my

Moroccan data, as in the following, where the morpheme can simultaneously be read as both a relativizer and purposive:

- (6) *dāru* 'īdrāb **bāš** mā-ynəqṣū-lhūm-š f 'ādād sā'āt
 did.3PL strike ? NEG-decrease-3PL.DAT-NEG PREP number hours
 l-'āml
 ART-work
 “they had a strike, by which [the admin] wouldn’t reduce their hours”
or “they had a strike in order that [the admin] wouldn’t reduce they their hours”

Because ambiguous tokens of *bāš* can be read as the historical “links” that connect earlier meanings to later ones, I use them as a proof of concept in suggesting that most likely historical development of the North African purposive is in the same order I gave the first three of its extant grammatical roles above: INTERROGATIVE → RELATIVIZER → PURPOSIVE. The data also indicates that the commonly cited purposive itself can give way to further extension: PURPOSIVE → COMPLEMENTIZER → NOMINALIZER. Taken together, we can thus envision a five-stage grammaticalization process, along which the familiar “purposive” meaning of *bāš* only represents a medial stage. I conclude the present study by making a case for some Moroccan varieties as having breached the final stage, and offer preliminary evidence that the commonly recognized purposive *bāš* has undergone further reanalysis in other North African varieties as well.

Universal Dependencies for Written Informal Arabic

Valeria Pettinari (Roma Tre University)

The present study describes the theoretic linguistic assumptions and methods employed in the creation of a Written Informal Arabic treebank, based on the Egyptian variety of Cairo, under the framework of the Universal Dependencies annotation scheme (UD). Traditionally, Arabic colloquial varieties are poorly described and barely considered in the field of Arabic Computational Linguistics and Natural Language Processing. When available, data from colloquial Arabic texts are variously described and analyzed, often considered as deeply divergent from standard Arabic structures. This work accounts both formal and informal expression of Arabic, that is the co-occurrence of formally standard and colloquial-oriented language, comprehensively defined as “Written Informal Arabic”.

The Universal Dependencies Written Informal Arabic Treebank (henceforth WIAUD) was created to (i) provide highly accurate linguistic data for which morphological, lexical and syntactic information is given, which follows a universal scheme of annotation in order to (ii) allow cross-linguistic investigation as well as (iii) effective comparison with standard Arabic, for which three UD treebanks are already found in the UD repository¹. WIAUD is based on texts drawn from the web: they show quasi-oral structures and are characterized by a high degree of orthographic variation and their language shows the co-occurrence of formal and informal language. A major issue is thus the consistent annotation of formal and informal linguistic structures in one treebank.

UD scheme was chosen because it provides a cross-linguistic morphological and syntactic annotation standard explicitly aiming at accounting for similarities and idiosyncrasies among typologically different languages.

In this work, major challenges in the annotation process are widely discussed, and they concern:

- Textual segmentation and sentence boundary identification;
- Orthographic variation;
- Tokenization and lemmatization;
- Classification of peculiar morphological and syntactic phenomena (discontinuous negation, light verbs and auxiliaries, modals, verbal aspect morphemes, pseudo-determines, final-clause *wh*- structures, adverbial structures etc., found in Egyptian Arabic);
- Lexical refinement of general Part-of-Speech classes;
- Functional syntactic annotation of ill-structured texts.

The risks of overgeneralization intrinsic to the cross-linguistic perspective underlying the annotation scheme have been effectively contained by means of a wide set of labels for morphological description refinement and the possibility to adapt the syntactic label to language-specific phenomena via sub-classification. Furthermore, the hybrid architecture of the UD scheme, envisaging both functional and structural syntactic analyses, ensures an adequate treatment of informal, ill-structured texts. WIAUD reveals interesting when it come to the possibility of comparative investigation with standard Arabic structures: not only does a comparative study offers a concrete opportunity to observe, on a wider scale, the emergence of structural similarities that can be eventually explained in a comprehensive manner but, more importantly, may be a first step towards the detection of structural similarities brought about by contextual contact and interaction between the formal and colloquial varieties found in written informal texts. Such perspective may provide a solid ground for the understanding of how the two influence each other, not only in mixed texts but also to a more general extent.

¹ Smrž et al., 2008. Taji et al., 2017.

Aspects of the Phonology and Morphology of Peninsular Dialects of Arabic (invited talk)

Stuart Davis (Indiana University)

This talk will bring together detailed description of various phonological, morphological, and morpho-phonological phenomena witnessed in different Peninsular varieties of Arabic based on research emanating from Indiana University and elsewhere. Some of the phenomena to be discussed are known, but perhaps the specific details are not, while other phenomena are less known. As an example of the latter, we will exemplify an unusual productive morphological augmentative witnessed in Ha'ili (northern Najdi) Arabic (Alashammari & Davis 2019) that includes forms like [k**la:b**] 'dog (augmentative)' for [kalb] 'dog' (where bold indicates pharyngealization), [ʃ**wɑ:ʕir**] 'poet (augmentative)' for [ʃa:ʕir] 'poet', and [s**ra:wi:l**] 'pants (augmentative)' for [sirwa:l] 'pants'. An analytical issue is whether these augmentatives are based on the noun forms shown or on the corresponding dialectal diminutives: [klejb], [ʃwejʕir], and [srejbwi:l], respectively. A morphophonological observation to be discussed concerns the realization of the 2nd person singular possessive suffix. In various Peninsular dialects the masculine possessive suffix will alternate between [-k], [-ka], and [-ak] but where the analysis indicates that /-k/ is the underlying form of the suffix. Compare, for example, the form [kutubk] 'your (m.sg.) books' in the Southwestern Saudi variety described by Alahmari (2018) with the Cairene form [kutubak] 'your (m.sg.) books' where both dialects allow for final clusters. The allomorphs [-ka] and [-ak] are exemplified in [malik-ka] 'your (m.sg.) king' and [baab-ak] 'your (m.sg.) door'. The analysis of this will be discussed. Furthermore, we will consider a variety of phonological issues that include how word-internal superheavy syllables (CVCC/CVVC) are avoided (or not avoided) in Peninsular varieties as exemplified by the difference between well-known Hijazi [baa.ba.na] 'our door' with low vowel epenthesis (where the period indicates a syllable boundary) from underlying /baab-na/ versus less known Hadhrami [baa.bna] (Bamakhramah 2009) where a potential superheavy syllable is avoided through onset maximization. We will also consider the effect of geminate consonants on word stress as observed by Watson (2002) for San'ani Arabic where a syllable closed by a geminate attracts stress even when there is another heavy syllable after it as in [dáv.wart] 'I/you (m.sg.) looked for' (versus [dawárt] in other varieties). The San'ani pattern seems to occur in other areas of the southwest part of the Arabian Peninsula. A somewhat different phenomenon that appears to be fairly widespread among Najdi varieties is final degemination (and stress shift) in bisyllabic words that end in an underlying geminate as illustrated by the difference between [ħagg] 'a truth' with [ʔil.ħag] 'the truth' with stress on the definite article or in the elative [ʔá.xaf] 'lighter' with initial stress where most dialects have [ʔa.xáff]. Finally, this talk tries to reach an understanding of some unusual features of Faiḫi Arabic including labiodentalization of some emphatics as in Upper Faiḫi Arabic [faħ.śa] 'he laughed' cognate with Classical Arabic [dʕaħaka] (Davis & Alfaifi 2019), the phonology of the [st] reflex of what is /sʕ/ in most varieties (Alfaifi & Davis to appear) and matters related to the distribution of the m-definite article in Faiḫi Arabic and in geographically proximate varieties (Alqahtani 2015).

**Language and Identity in the pre-Islamic Arabic of the Levant and North Arabia
(invited talk)**

Ahmad Al-Jallad (Ohio State University)

How did pre-Islamic speakers of Arabic self-identify? Was language an important feature of group identity? Did the ethnonym “Arab” always overlap with “Arabic speaking”? This talk will discuss these questions based on the epigraphic record and our increasing understanding of Old (pre-Islamic) Arabic based on documentary sources.

Kalām al-‘arab. Sources, Problems and Methods

Francesco Grande (University of Turin)

The study of the earliest stages of a language usually relies upon epigraphic and literary sources. The resort to either source, or both, crucially depends on their availability. Regarding Arabic, this kind of study focuses on materials from the pre-Islamic period, for which the availability of sources is unbalanced. The study of the earliest stages of Arabic has benefitted from epigraphic sources, made increasingly handy by modern archaeology, but still suffers from unavailability of literary sources¹. The gist of the problem is that several literary sources do record pre-Islamic linguistic materials, which they refer to as *kalām al-‘arab*², but are unavailable because of interpolation or inaccessibility. Interpolation mainly affects poetical literature, resulting in unreliability of poetic *kalām al-‘arab* (also known as ‘pre-Islamic’ poetry). Inaccessibility is typical of technical literature (e.g., lexicography), where records of ordinary *kalām al-‘arab* are generally incidental and scattered across grammatical or lexicographical discussions³.

This talk presents a research line aiming at increasing the availability of literary sources and, relatedly, of materials from ordinary *kalām al-‘arab*, to integrate them in the study of the earliest stages of Arabic. In light of the above, increasing the availability of these sources and materials implies: (i) improving accessibility to lexicographical literature and its materials from ordinary *kalām al-‘arab* (ii) ensuring their reliability. The methodological core of the research is the notion of protolanguage, intended as any language characterized by scarcity of written records⁴. The protolanguage has two defining features:

- consists at least of words with basic meanings (Grundwortschatz)
- has internal variation

The next methodological step is interpreting ordinary *kalām al-‘arab* as a protolanguage, because of the scarcity of its written records. This defines the research scope as follows. First, since ordinary *kalām al-‘arab* consists (at least) of words with basic meanings, the researcher can cluster these meanings into basic semantic fields, which are a valid word retrieval criterion in lexicographical works: many early Arabic dictionaries (about 800 CE) were organized precisely by semantic field (*mubawwab*)⁵. Once retrieved the desired semantic field in early *mubawwabs*, the researcher narrows the search down to a word of ordinary *kalām al-‘arab* conveying the desired basic meaning. The Grundwortschatz feature of protolanguage therefore serves the aim of (i). This talk also discusses which version of Grundwortschatz and which *mubawwabs* to use concretely to describe ordinary *kalām al-‘arab*. Second, since ordinary *kalām al-‘arab* has internal variation, for any word retrieved along the above lines the researcher collects and lists all variants recorded in early *mubawwabs*. The variationist feature of protolanguage therefore serves the aim of (ii): the more variants a retrieved word of ordinary *kalām al-‘arab* possesses, the more reliable it is.

With the same aim, the researcher privileges those variants recorded in early *mubawwabs* with explicit mention of original informants, through formulas like *qāla(t) X*, with X a tribe or a tribe member.

This talk discusses the word *‘akala* to illustrate this research line step by step.

¹ Al-Jallad, Ahmad. 2018. The Earliest Stages of Arabic and Its Linguistic Classification. In *The Routledge Handbook of Arabic Linguistics*, edited by Elabbas Benmamoun and Reem Bassiouney. New York: Routledge: 315-331. Al-Jallad, Ahmad. 2018. What is Ancient North Arabian? In *Re-Engaging Comparative Semitic and Arabic Studies*, edited by Daniel Birnstiel and Na’ama Pat-El. Wiesbaden: Harrassowitz: 1-45.

² Baalbaki, Ramzi. 2014. *The Arabic Lexicographical Tradition*. Leiden: Brill.

³ Renan, Ernest. 1855. *Histoire générale et système comparé des langues sémitiques*. Première Partie: Histoire générale des langues sémitiques. Paris: Imprimerie Impériale.

⁴ Tovar, Antonio. 1954. Linguistics and Prehistory. *Word* 10: 333-350.

⁵ Baalbaki, Ramzi. 2014. *The Arabic Lexicographical Tradition*. Leiden: Brill

A historical scenario fro the distribution of Tanwīn Alif in Christian Palestinian Arabic texts

Phillip W. Stokes (The University of Tennessee)

Early Islamic texts produced by Christians in south Palestine are widely accepted as among the earliest examples of Middle Arabic. The feature perhaps most characteristic of this label is the non-Classical distribution of tanwīn alif. Most scholars of Christian Palestinian Arabic texts (e.g., Knutsson 1974; Bengtsson 1995; Hjälms 2016) have followed Blau (1967: §215-228) in analyzing these as reflecting hyper-corrections based on an imperfect grasp of Classical case inflection.

The assumption that early Islamic Arabic compositions derived from Classical Arabic, however, is now clearly untenable. Pre-Islamic Arabic was clearly varied and diverse (AlJallad 2018). Further, Al-Jallad (forthcoming) argues convincingly that the writing tradition behind, e.g., early Papyri, as well as the Violet Psalm Fragment, reflects a non-Classical Ḥiḡāzī one. The Qur'ān itself reflects this background (cf. Van Putten and Stokes 2018) In this paper, I argue for a different framework for understanding this feature, and the nature of these texts. The non-Classical roles of tanwīn alif are regular, and I propose are explainable simply as examples of dialectal tanwīn. I will argue that the oddities of distribution result from the use of an inherited Ḥiḡāzī writing tradition, which included a functional accusative written with alif, to write varieties that possessed forms of dialectal tanwīn. Some authors utilized the Ḥiḡāzī system exclusively, while others combined and mixed the two in various, creative and idiosyncratic ways.

Non-Standard Written Practices in Andalusī and Maghrebi Medieval Texts

Estefania Valenzuela Mochon (The University of Texas at Austin)

The common practice of “correcting” Arabic manuscripts prior publication to conform to the stylistic norms of the standard variety, since late 19th-early 20th century, has obscured much of the history of written Arabic (Lentin 2011; Brustad 2017). Moreover, the tradition of adapting the language of the manuscripts to meet a specific criterion is believed to have contributed to “the erasure of language variation that might have existed in pre-modern times” (Brustad 2017, 44).

In this study, I explore a methodology grounded on a close rereading of early sources as an alternative approach to shed some light into the history of written Arabic in the Islamic West. More specifically, I focus on the analysis of 7 manuscripts that deal with the topic of *ḥisba*, or urban market regulations. These texts are copies of works originally composed between the 10th and 13th centuries CE by three Andalusī authors: ‘Abd al-Ra’ūf, Ibn ‘Abdūn, and al-Saqāfī. In addition to offering vivid depictions of the local economies in important trade centers of al-Andalus, the value of these documents lies in their status as “layered texts.” That is, the *ḥisba* manuscripts reveal a language that encompasses the use of classical, non-standard, and local features. The alternation between linguistic forms is understood here as a pragmatic strategy that allows the speaker/writer to convey specific meanings.

In order to illustrate the type of language variation found in my corpus, I discuss two distinctive functions related to the grammatical category of mood identified in the Andalusī books: the usage of the verbal ending *-ū* with verbs of duty and obligation (1-2) and *-ūna* as a marker of future tentative actions (3). In the following examples, (a) shows the standard norm, while (b) represents the forms found in the *ḥisba* manuals:

1. a) *yu ‘marūna* ‘an lā yabī‘ū *ḍahab* *bi-fiḍḍa*
order-IND.3PL that not sell-SBJV.3PL gold for-silver
 b) *yu ‘marū* ‘an lā yabī‘ū *ḍahab* *bi-fiḍḍa*
order-JUSS.3PL that not sell-SBJV.3PL gold for-silver

They [money-changers] shall be ordered not to sell gold for silver.

2. a) *yu ‘addabūna* *al-šuyūḥ* ‘alā *ḍalika*
punish-IND.3PL the-sheiks on that
 b) *yu ‘addabū* *al-šuyūḥ* ‘alā *ḍalika*
punish-JUSS.3PL the-sheiks on that

They shall be disciplined by the sheiks for that.

3. a) *yumna ‘ūna* ‘an lā yantafi‘ū *bi-tiyāb* *al-ḥām...*
 prevent-IND.3PL that not use-SBJV.3PL with clothes the-cotton-raw
tumma yuqašširū-hā *ba‘da ḍalika*
 then **whiten-SBJV.3PL-it-F.SG** after that

- b) *yumna ‘ūna* ‘an lā yantafi‘ū *bi-tiyāb* *al-ḥām...*
 prevent-IND.3PL that not use-SBJV.3PL with clothes the-cotton-raw
tumma yuqašširūna-hā *ba‘da ḍalika*
 then **whiten-IND.3PL-it-F.SG** after that

They [fabric sellers] should be prevented from using clothes made of raw cotton... to then whiten them.

In examples (1) and (2), the verbs *ya'mur* 'to order' and *yu'addib* 'to discipline' take the *-ū* inflection as they introduce a new sentence, a context that requires the indicative form *-ūna* according to the prescriptive norm. The semantic property of obligatoriness embedded in the lexical meaning of these verbs seems to be the force driving the use of the *-ū* inflection in these cases. On the other hand, example (3) shows a construction with the particle *tumma* 'then' and the verb *yuqaṣṣir* 'to whiten,' in which the verbal ending *-ūna* appears to attach an intensitive modal value to the subordinated verb.

Overall, the examination of the language in these manuscripts suggests that written practices in the western Islamic world did not always adhere to the of rules of Classical Arabic. This finding is consistent with similar results documented in other early and medieval Arabic written texts from the East (Blau 1966, 2002; Hopkins 1984; Lentin 1997), which further confirms the existence of a rich tradition of variation in Arabic writing forms and, most importantly, it proves the transmission of this linguistic diversity to the western regions of the Arabic speaking world.

The role of education and dialect contact in the realization of (ð^s) and (d^s) in a speech community of Palestinian refugees in Syria

Ourooba Shetewi (Newcastle University)

A merger between (ð^s) and (d^s) occurs in the vast majority of Arabic dialects (Al-Wer 2003). A split in their realization is only maintained in Standard Arabic and a handful of isolated South Western Saudi Arabian dialects (Watson & Al-Azraqi 2011). For example, both variables are realized as [ð^s] in Bedouin dialects and some rural dialects, whereas they are both realized as [d^s] in urban dialects. [z^s] also occurs as a lexically-conditioned realization in urban dialects. Given the overlap in their vernacular and standard realizations in different speech communities, an examination of speakers' use of these variables gives valuable insights into their sociolinguistic awareness. Therefore, this paper discusses the role of education and dialect contact in the realization of (ð^s) and (d^s) by children and adolescents in a Bedouin speech community of Palestinian refugees in Syria. The two variables are realized as [ð^s] in the traditional dialect of the community under study. However, exposure to [d^s] occurs through contact with the urban dialect of Damascus, where both variables are realized as [d^s], and exposure to SA through education where a split in their realization is maintained.

Data presented in this paper is part of a larger research project on acquisition of sociolinguistic variation in dialect contact situations. Two female fieldworkers, a local and an urban speaker, conducted sociolinguistic interviews and play sessions with 40 girls and boys (3;7- 17;9) in order to elicit spontaneous speech and examine accommodation patterns across different interlocutors. A picture-naming task was also carried out by the local speaker to examine the emergence and patterns of register variation in the speech of participants.

Distributional analysis of the variables and an overview of their realizations across different contexts revealed intriguing patterns of their variation. While the local variant [ð^s] was the majority variant for both variables, its use as a realization of (ð^s) was noticeably higher (72.5%) than in the case of (d^s) at 54%. Furthermore, a paired sample t-test that examined the influence of context and interlocutor on speakers' use of the variables revealed that both had a significant effect. For example, use of the urban variant [d^s] as a realization of both variables was significantly higher in the interview with the urban fieldworker than in the interview with the local interviewer (figures 1 & 2). A significant difference in the realization of (d^s) as a function of context was also revealed with use of the standard variant in the picture task being significantly higher by comparison to the interview context (figure 3).

Results on the variation of these two variables show that children and adolescents have an awareness of the standard split between the variables as well as their status in different spoken varieties despite the merger in their own dialect. This suggests that both education and dialect contact play a role in the linguistic behavior of young speakers in the community.

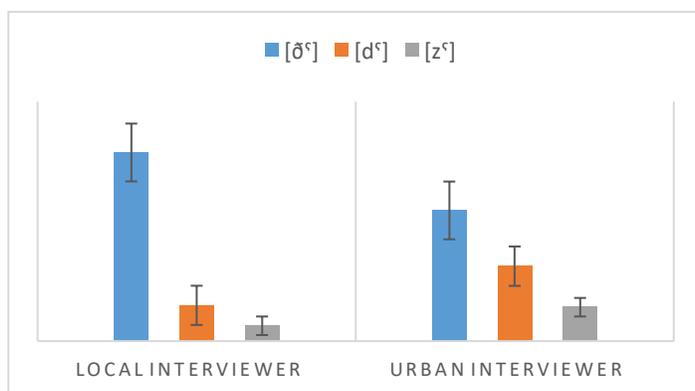


Figure 1 Distribution of (d^h) variants across interview contexts

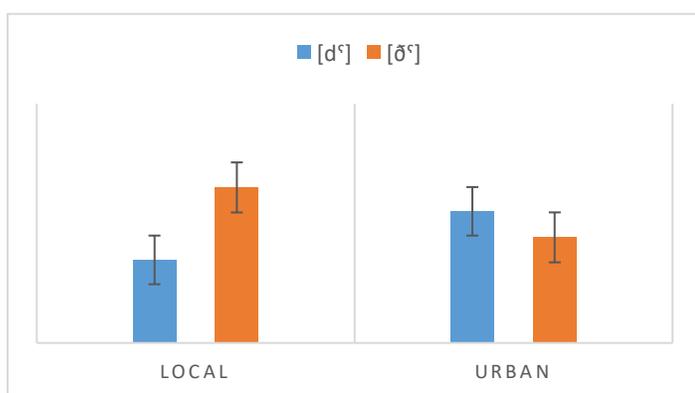


Figure 2 Distribution of (d^h) variants across interview contexts

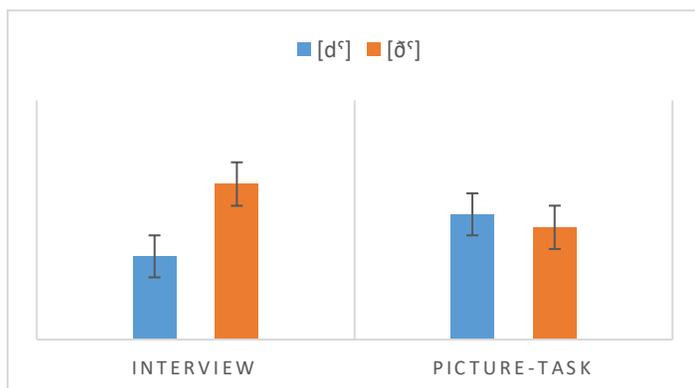


Figure 3 Distribution of (d^h) across contexts

The Semantics of Substrate Lexica in Colloquial Levantine Arabic: Dynamics of Survival and Naturalisation

Mila Neishtadt (University of Cambridge)

A *language shift* is a process whereby a speech community ceases to speak one language and acquires another one instead. Features of the abandoned language that are extant in the supplanting language after the language shift may be regarded as *substrate interference* or *substrate influence*. The influence of *substrate language* on the *target language* (hereafter TL), may be traced in several aspects of linguistics: phonology, morphology, syntax, lexicon, etc.

The common explanation for substrate influence on the lexical component of the TL is usually sought in the existence of *lexical gaps* between the *substrate language* and the TL in the semantic fields of toponyms, topography, and agriculture that are unique to the geographical region in question (Thomason and Kaufman 1988). Accordingly, one might expect that the vast majority of *substrate words* would be nouns or nominal forms.

My study of Aramaic substrate lexica in Syro-Palestinian colloquial Arabic is based on two main types of sources: a) basic sources for compilation of a preliminary non-sifted list of potential *substrate words*: Syro-Palestinian Arabic dialectal dictionaries, ethnographic sources, previous research; and b) reference sources for identification of these potential substrate words: Arabic and Aramaic lexicography as well as Rabbinic Hebrew literary sources. The suggested methodological framework of the study is a solid set of linguistic criteria for the process of identification of substrate words: (a) Phonology (b) Morphology (c) Attestation and distribution in Aramaic and Arabic (d) Semantics (for details, see: Neishtadt 2015). As a result of this investigation, I identified 435 substrate lexical entries (derived from ~100 etymons).

The semantic analysis of the above mentioned substrate vocabulary suggests a more complex perspective that sheds light on both the particular case of substrate influence on the lexica of Syro-Palestinian colloquial Arabic as on the general theory. As to the division of the analysed substrate words into parts of speech, we found that 63% of the examined lexical entries were nouns, 35% – verbs and 2% – adjectives. The leading semantic field is agriculture and adjacent domains: yoke and plough, agricultural tools and implements, field crops (olives, grapes, figs), climate and seasons of the year, soil types, domestic animals and animal breeding, flora and fauna. However, many substrate words in Levantine colloquial Arabic have no connection to agriculture or related semantic realms. Among these words we find: strong language and terms of endearment (nouns and adjectives); movement and intense motion, beating and physical injury, pejorative terms to describe actions or situations (verbs).

In the proposed talk I shall present and analyse the gathered data and suggest several possible explanations of the findings based on a) The specific nature of the Arabisation process of the area in question b) Identification of possible universal tendencies which may suggest improvement of the general theory of substrate influence.

A sketch of the Bedouin dialect of the Shahibaat (Egypt)

Valentina Serreli (University of Bayreuth)

The dialect of the Shahibaat, a branch of the Awlaad ʕAli tribe of the North Western Egyptian coast, which settled in the Siwa Oasis at the beginning of the 20th century (Bliss 1981: 61), has never been studied. The dialect of the Awlaad ʕAli is partially described in Maṭar (1967) and in some notes accompanying four texts published in Behnstedt & Woidich's Egyptian dialect atlas (1985, Band 3: 244-263). This talk offers a sketch of the Shahibaat dialect, based on four interviews conducted with Shahibaat women and men in 2019/2020.

The dialect of the Awlaad ʕAli (Behnstedt & Woidich 1985) shares a number of traits with the Eastern Libyan dialects (Owens 1983, 1984; Benkato 2014). This study shows that the Shahibaat dialect also displays such features and belongs to this group. Besides those traits shared by virtually all the Maghribi Bedouin Dialects, e.g. the paradigm *niktib-nikitbu* of the 1SG and 1PL of the imperfect verb and the voiced realization of *q as [g] (1), the Shahibaat dialect also displays features proper of the easternmost Maghribi dialects, as for instance the interdental fricatives [t̪] [d̪] [ḏ̪], with merger of [d̪] and [ḏ̪] into [ḏ̪] (1); short vowels in open syllable; *ghawa*-syndrome; and the gender distinction in the 2P and 3P plural in verbs (2a, 2b) and pronouns (2c, 2d). And it also displays Egyptian Arabic lexical influence.

(1) *We bring fabric / from outside white fabric and from inside red fabric*

n-žiib-u	gmaaš / min	barra	gmaaš	ʔabyaḏ u	min	žowwa	gmaaš	ḥamar
1-bring.IPFV-PL	fabric / from	outside	fabric	white	and from inside	fabric	red	

(2a) *Five women come*

i-ž-an	xamas	ṣubaaya
3-come.IPFV-PL.F	five	woman.PL

(2b) *They (m) give the wool to the women*

yi-ṣaṭ-u	ṣ=ṣoof	li=ṣ=ṣubaaya
3-give.IPFV-PL.M	DEF=wool	to=DEF=woman.PL

(2c) *To the girls and their (f) mother*

li=l=banawit	u	ʔumm- hən
to=DEF=girl.PL	and	mother- 3PL.F

(2d) *He loaded them (m) on his camel*

rakkab= hum	ʕala	žmaal=a
load.PFV= 3PL.M	on	camel=3SG.M.POSS

The talk contributes to the field of Arabic dialectology as it proposes data on an under described Arabic Bedouin dialect. Moreover, it intends to show whether and how the distance from the Awlaad Ali of the coastal cities and the over one century-long exposure to the other languages spoken in the Siwa Oasis, i.e. Siwi Berber and Egyptian Arabic, affected it.

Toward a reassessment of Tunisian village dialects: the case of Chebba

Luca D'Anna (University of Naples "L'Orientale")

The traditional classification of Tunisian dialects, first proposed by Marçais (1950), distinguishes three categories: Bedouin (Hilali), sedentary (pre-Hilali) and rural / village dialects. While the existence of first two categories is supported by reliable isoglosses, village dialects are characterized by a mixture of Bedouin and sedentary traits, and their existence as an independent group was recently questioned by Mion (2015; 2018) and Guerrero (2018). This paper presents preliminary data on the Arabic dialect of Chebba (Tunisia), based on fieldwork conducted in February 2020. The dialect of Chebba, on the Tunisian Sahel, represents a mixed variety, apparently originated out of contact between Hilali and pre-Hilali varieties spoken in the area. It features Hilali phonology but, at the same time, a markedly pre-Hilali morphology. One neighborhood of the town, which was a separate village until the second half of the 19th century, still bears traces of its original Hilali morphology, which appears to be losing ground to pre-Hilali forms employed in the other neighborhoods. Measured against Marçais' criteria, the Arabic of Chebba might be considered as a rural variety. The paper, however, advances the hypothesis that the category of Tunisian 'village' dialects might be better explained in terms of contact or mixed varieties, rather than as a self-standing class. The explanation of the peculiar mixture of Bedouin and sedentary traits of such varieties requires a detailed analysis of the historical and sociolinguistic processes leading to their formation, which are here exemplified using Chebba as a case study.

Parallel grammars in one language: An analysis of Arabic loanwords in Saraiki

Nasir A. Rizvi (Lasbela University), Shah Bibi (Lasbela University), Tooba Sahar (Lasbela University)

This paper highlights changes which occur in Arabic words taken as loans in Saraiki. Saraiki is an Indo-Aryan language of Pakistan (Bashir and Connors, 2019). It is a quantity sensitive language, which always stresses the heaviest syllable in a word. If a word has two syllables of equal weight, stress falls on the left (which we call left-headedness). Saraiki does not allow an unstressed heavy syllable on left-edge of a word. These constraints trigger paradoxical changes, which result into insertion or deletion in hundreds of Arabic loanwords in Saraiki.

The words of LH (Light-Heavy) syllables in Arabic change into **HH** in Saraiki (Stressed syllables are highlighted bold). This is done by insertion of a consonant in the light syllable, e.g. Arabic word /na.qi:/, which has CV.CVV (Light-Heavy) syllables, is adapted as /**nəq**.qi:/ (CVC.CVV) in Saraiki. (see other examples in Appendix-A). Contrary to this, sometimes Saraiki speakers delete a consonant or a vowel to satisfy the left-headedness constraint which does not allow an unstressed heavy syllable on the left edge of a word. For example, Arabic words like /fa:.**ru**:q/ (CVV.CVVC) become /fə.ru:q/ (CV.CVVC) in Saraiki to resolve the conflict between the ban an unstressed heavy syllable and quantity sensitivity. So, Saraiki reduces a mora of the left heavy syllables of such words when the right syllable is super-heavy. The result is vowel shortening in the penultimate syllables of such bi-syllabic words. (See other examples in Appendix-B). Similar deletion also targets consonants resulting into degamination. For example, the Arabic words like /qah.ha:r/ change into [qə.ha:r] to satisfy the same constraints. (Appendix C for more data).

An important thing, in this regard, is that Saraiki accepts **LH** type of indigenous words which bear stress on ultimate syllable, but it does not accept the same structure in Arabic loanwords. Being quantity sensitive, Saraiki stresses heavy ultimate syllables in a word of light-heavy (**LH**) syllables (see other similar indigenous Saraiki words in Appendix-D). This confirms that Saraiki is not strictly a Trochaic language. It allows more marked Iambic foot form. However, this structure is only allowed in indigenous Saraiki words but not in loanwords. (see Appendix-A for examples of loanwords which are not accepted in iambic forms). In other words, Saraiki obeys markedness constraints in loanword grammar but in the L1 grammar, it is more faithful to the original input. Similar trend is also observed in another class of Arabic loanwords. Saraiki does not accept clusters of coronal+[r] (DR clusters) although similar structures do exist in indigenous Saraiki words. For example, in Saraiki the word like ‘gadr’ (spot) is legitimate (see other examples in Appendix E) but similar Arabic word like ‘badr’ is not acceptable and is changed into ‘bəd.dər’ in Saraiki by insertion of a vowel and a consonant in the original Arabic word. Significantly, the epenthetic vowels and consonants in the adapted loanwords are harmonious with the stem vowels and consonantly respectively, as the examples in Appendix F show.

“Face” in the Male World of Tribal Reconciliation (atwa) in Jordan

Abeer Malkawi (University of Brighton)

My paper examines the ritual practices of atwa in Jordan with pragmatic methods. Atwa, عظة in Arabic, is a ritual of tribal reconciliation occurring in parallel to the Jordanian state legal tribunals. The function of atwa is to establish canals of communication between the victim’s clan and the offender’s clan and to prevent acts of retaliation. Atwas are performed and resolved amongst men but –recently and thanks to modern technology- many cases have been uploaded on YouTube. I collected atwa cases via the YouTube search engine between 2013 and 2017. The collected cases deal with fights, fatal car accidents and murder.

In my analysis of face in the atwa cases, I will elaborate on the Brown and Levinson (1978/1987) positive vs negative face framework and build on Spencer-Oatey’s framework (2000) to highlight some transferrable values such as respect of tradition, kindness, forgiveness in the Jordanian context. I will then analyse the specifics of ‘collective’ face in the atwa cases with a focus on the social roles at play, for example when - the speaker on behalf of the delegation is also an important member of Jordanian society. Other examples of ‘collective’ face show references to holy books, the Quran, the Bible and the New Testament, which are holy books in the Jordanian context and therefore features of face for both the delegation representing the offender and the victim’s clan. My case studies will establish the limits of the positive/negative face framework and, - building on the Spencer-Oatey’s framework-, will demonstrate that larger social interactions are at play and need to be evaluated when analysing face.

Pranks as a new eliciting technique for the study of lexical diffusion in Moroccan Arabic

Jacopo Falchetta (IREMAM)

The variation between diachronic reflexes of etymological /q/ has been investigated by a great number of Arabic dialectologists (Abdel Jawad 1986; Al-Khatib 1988; Al-Wer 1999, 2007; Al-Wer & Hérin 2011; Amara 2005; Gibson 2013; Habib 2010; Haeri 1996; Jabeur 1987; Taine-Cheikh 2000 among others), including those working on Moroccan varieties (particularly Benthami 2007; Hachimi eg. 2007, 2011; Heath 2002; Moumine 1990). When they have not based their analysis on second-hand data, these authors have usually made use of “classic” data elicitation techniques, including questionnaires, face-to-face and group interviews, recording of spontaneous conversations, reading passages, word lists and ethnographic observations. The effectiveness of these techniques for the purposes of dialectological and sociolinguistic research has been demonstrated in the literature by the quantity and quality of data that they have allowed to obtain, and by the important advancements that have been consequently made in the discipline. However, they present fundamental problems when they are aimed at eliciting specific variables, as they either provide the researcher with great amounts of redundant data (as in the case of interviews), or cause the informant to over-focus on the variables sought, thus risking to obtain non-spontaneous answers (as in the case of questionnaires or word lists). In a recent study of variation in Moroccan Arabic, I examined the alternation between /g/ and /q/ as a case of lexical diffusion, and chose to focus on a limited set of lexemes that appeared to be alternatively assigned to both the /g/ and /q/ phonological classes by members of the community studied (the town of Temara, Morocco): an example is their alternating between /wgəf/ and /wqəf/, “to stop”. In order to obtain satisfying amounts of occurrences of these particular lexemes, I elaborated a specific type of test consisting in showing a speaker a hidden-camera video, requesting him/her to tell it to another speaker (who has not watched the video) and then asking the latter questions about the narration he/she has just listened to. The pranks are chosen according to whether they are likely to cause the speaker viewing the video to utter the targeted lexemes when describing the prank to his interlocutor (e.g.: a hidden camera in which the victim is forced to stop at a red traffic light is more likely to elicit the lexeme /wgəf/, or /wqəf/), or whether they allow the researcher to indirectly elicit the same lexemes from the other speaker when asking the questions aimed at checking his/her having understood what happens in the prank (e.g.: “what does the victim do when he gets to the traffic light?”). Using hidden-camera pranks in order to elicit data is an attempt to overcome the shortcomings of the techniques mentioned above, as well as to obtain cross-comparable data from a relatively large number of speakers. In the present paper, I shall expose in detail how the test was concretely administered to my informants, and discuss how the results contributed to the aims of the research and what proved to be the advantages and the shortcomings of this technique.

fuṣḥá, ‘āmmīyah, or both?: a new theoretical framework for contemporary Arabic writing

Saussan Khalil (University of Cambridge)

In the past few years, interest in colloquial or mixed Arabic writing has (re)emerged. Doss & Davies's (2013) compilation of colloquial Egyptian Arabic writing since the Middle Ages has provided a much-needed historical overview of what at times appears to be a modern phenomenon.

Several studies such as Rosenbaum (2000), Ramsay (2012), Kosoff (2014), and Håland (2017) have identified the use of colloquial dialects, often mixed with *fuṣḥá*, in contemporary writing both online and in print. This in addition to the study of Middle Arabic (Bellem & Smith, 2014) as well as 'mixed' styles (Mejdell, 2006 and 2011-12) in Arabic writing.

These examples of contemporary mixed or colloquial Arabic writing provide useful observations about the current state of Arabic language usage, since previously observations around mixing were focused mostly on speaking, rather than writing such as Eid (1988), Bassiouney (2006) and Mejdell (2006).

However, while some observations about colloquial and mixed Arabic writing have been made, it is critical to try and bring these and others together in an attempt to provide a theoretical framework for Arabic writing, much like Badawi's (1973) study of contemporary Arabic in Egypt at that time. Høigilt, and Mejdell (eds.) (2017) go some way in explaining the current writing phenomenon through the lens of political change.

This study brings together original analysis of contemporary colloquial and mixed texts, as well as a review of the studies above, and proposes a new theoretical framework for Arabic writing, that takes into account writing practices in Classical, Modern Standard, Colloquial and Mixed Arabic writing. It identifies seven distinct writing 'styles' that can be used consecutively or simultaneously by the same writer in the same or different texts.

It further includes examples from different time periods, genres, political contexts and motivations and suggests that viewing these all together, rather than separately, can go some way towards providing an overarching framework that is flexible and adaptable, and explains the various phenomena observed in Arabic writing over different time periods, including the current one.

Lexical changes in Arabic newspaper writings: A corpus-based comparison of 2018 and 1950 Arabic newspapers

Sarah Alajlan (University of Leeds) & Claire Brierley (University of Leeds)

Studying language change is one of the most challenging topics in the field of corpus linguistics. The challenge stems from two facts. First, there is a shortage of ancient texts in a machine-readable format, and second there is the difficulty of identifying linguistic changes electronically across two or more corpora. This paper attempts to shed light on lexical changes that have happened in Arabic newspaper writings by comparing two sub-corpora that have been built specifically for this study (Author, 2019). It is part of an ongoing PhD study that uses automatic analysis followed by manual investigation of the lexicon and the syntax of Arabic newspaper writings.

The corpus texts were collected from three Arabic newspapers published in two Arabic countries. Al-Ahram was chosen to represent Arabic newspaper writing in Egypt in 1950 and in 2018. Alriyadh was chosen to represent Arabic newspaper writing in Saudi Arabia in 2018. Because Alriyadh was not published back in 1950, Umm Al-Qurā was selected to represent Arabic newspaper writing in 1950.

New methodology has been presented for the purpose of identifying lexical changes. At the beginning of the analysis, word frequency lists were generated using the Sketch Engine toolkit in order to obtain a clear picture of the nature of linguistic data in the corpus. Six lists were generated for each sub-corpus: (1) the entire 1950 sub-corpus list; (2) the entire 2018 subcorpus list; (3) the sub-corpus of Al-Ahram 1950 list; (4) the sub-corpus of Umm Al-Qurā 1950 list; (5) the sub-corpus of Al-Ahram 2018; and (6) the sub-corpus of Alriyadh 2018.

Since the lists were too long (approximately between 1,800 words and 8,000), manual analysis was not a good option. Instead, analysis of word lists was done via a Python script. The `set.intersection()` method was used to return the set of shared words between any two input lists, and the `set.difference()` method was used to identify the set of words unique to each of two input lists. Three return lists were generated for each comparison: the entire 1950 versus the entire 2018 corpus; the 1950 versus the 2018 Al-Ahram sub-corpus; and the 1950 Umm Al-Qurā versus the 2018 Alriyadh sub-corpus.

The results show that there have been visible changes in the lexicon of Arabic newspapers. The changes were classified into five categories: (1) increase or decrease in the frequency of occurrence; (2) semantic changes; (3) changes in form; (4) new emergence; and (5) disappearance of lexical items. Automatic extraction of frequent words and identifying shared and unique words in both sub-corpora was not sufficient for identifying the semantic changes especially in the intersection lists; manual investigation was needed to highlight some changes that happened. The manual analysis reveals some linguistic trends that were noticed in semantics of some lexical items. The phenomenon of “monosemisation” was noticed in this study where semantic shift has happened with respect to some polysemic terms such that they have become monosemic.

Corpus-based Grammatical Investigations in Media Arabic

Ashraf Abdou (University of Oxford)

Corpora have made many significant contributions to the study of grammar. They have been very useful particularly in identifying and examining issues such as frequent and infrequent grammatical structures, meaning-pattern relationships, pragmatic functions of grammatical constructions, grammatical features of the co-text, and grammar-lexicon relationships. However, so far little attention has been paid to the role that corpora can play in the study of Arabic grammar. This work takes a step to fill this gap by shedding some light on how corpora can help in this area, particularly with respect to media Arabic.

It mainly focuses on investigating two issues, first, the grammatical polarity of the Arabic exceptive particle *siwaa*, and, second, the lexical meanings typically communicated by single words functioning as adverbs of manner *aHwaal mufradah*.

In order to answer its research questions, the study makes use of some available corpora of Modern Standard Arabic, e.g. relevant sections in the International Corpus of Arabic, a grammatically annotated corpus that has been under development in Bibliotheca Alexandrina, and the King Abdulaziz City for Science and Technology (KACST) Arabic corpus.

With regard to the exceptive particle *siwaa*, preliminary findings show that it is mainly a negative polarity item. Its corpus examples typically contain explicit negative particles. But, sometimes the negation is understood from the context, as in the case of certain kinds of rhetorical questions.

As for the meanings conveyed by adverbs of manner, the analysis of the 316 examples of *al-Haal almufradah* in a corpus of news articles shows that about 55% of the data examples denote meanings that relate to ‘verbal expression’, while about 15% of the examples denote the meanings of ‘togetherness’ or ‘generality’.

Finally, the paper briefly discusses some implications of corpus-based grammatical studies in the area of Arabic language pedagogy.

Functions and Collocations of “ʔeeh” in Egyptian TV Drama: A Corpus-based Study

Mona Azam (The American University in Cairo)

*ʔeeh*¹ “what” is a frequent word in Egyptian colloquial Arabic (ECA), and it is often taught to learners of Arabic as a foreign language at the elementary level, however very few materials if any discuss the non-interrogative or pragmatic functions of *ʔeeh*. The aim of this corpus-based study is to investigate pragmatic functions and collocational behaviors of *ʔeeh* in the language of Egyptian television drama. The current study is based on an ECA written corpus consisting of transcriptions of 14 episodes of two Egyptian family TV series that were broadcasted in 2017 namely *sa:biʕ ga:r*² “The seventh neighbor” and *abu: el-ʕaru:sa*³ “The bride’s father”.

Wordsmith Tools software was used to analyze the data. Concordance lines containing *ʔeeh* were sorted by the researcher into different categories representing the different functions. A list of functions was adopted from Fareh and Moussa’s (2008) corpus-based study titled “Pragmatic Functions of Interrogative Sentences in English: A Corpus-based Study”. The analysis showed that *ʔeeh* has 21 functions in the current corpus (18 functions from Fareh and Moussa’s (2008) list and three new ones) including seeking clarification or information, showing irritation, showing admiration, showing satisfaction and dissatisfaction, drawing the hearer’s attention, negation, and disapproval. Moreover, results showed that *ʔeeh* is associated with a number of collocates to represent some of these functions such as *baʔu:llak*, *bass*, *walla:*, and *fi:h*⁴. These results offer an excellent opportunity to increase the accuracy of Arabic dictionaries and the pragmatic awareness of AFL learners. Consequently, AFL learners can become pragmatically appropriate with their language use and avoid the risk of sounding rude, disrespectful or offensive (Reynolds, 1991).

¹ إيه

² سابغ جار

³ أبو العروسة

⁴ بقو لك, بس, وللا, فيه

The Role of Multilingual Patterns and Functions on Branding and Self Branding: An Investigation of Practices by Saudi Users on Twitter

Reem Al Madani (Cardiff University)

The online presence of corporations and social media influencers plays a significant role in the way they communicate on the various social media platforms. The increasing number of internet users means that corporations and social media influencers build creative social media strategies and interactive posts in order to engage with their followers and encourage them to share the content related to them and their brand. The current study examines the digital communication of corporations and social media influencers with a focus on the multilingual practices used to communicate with the existing and potential virtual audience on Twitter in Saudi Arabia. The study aims at investigating the use of the different forms of Modern Standard Arabic, Colloquial Arabic and English, in addition to the use of punctuation to express emotions and emoji in more than 13,000 tweets from corporations' and social media influencers.

The study combines the observation of online discourse and contact with its social actors. The presentation will discuss the findings of the analysis of a dataset which was collected from the official Twitter accounts of 50 corporations and 30 social media influencers. It will show the different multilingual patterns used by the different corporations and the social media influencers. In addition, it will look at the different functions of the multilingual tweets that emerged in the data in relation to previous studies such as Androutsopoulos (2013) and Halim & Maros (2014). The corporations' accounts represent different businesses such as oil industries, retail, sports, automotive, and education. On the other hand, the social media influencers represent different interests such as sports, media, technology, and science. Furthermore, the presentation will give an overview of the interview results with a sample of marketing executives from the corporations and social media influencers to reveal the motivations for the multilingual practices and its relation to the branding of the corporations and the self-branding of the social media influencers.

CEFR Classified Arabic Vocabulary List

Nouran Khallaf (University of Leeds), Serge Sharoff (University of Leeds) & Micheal Ingleby (University of Leeds)

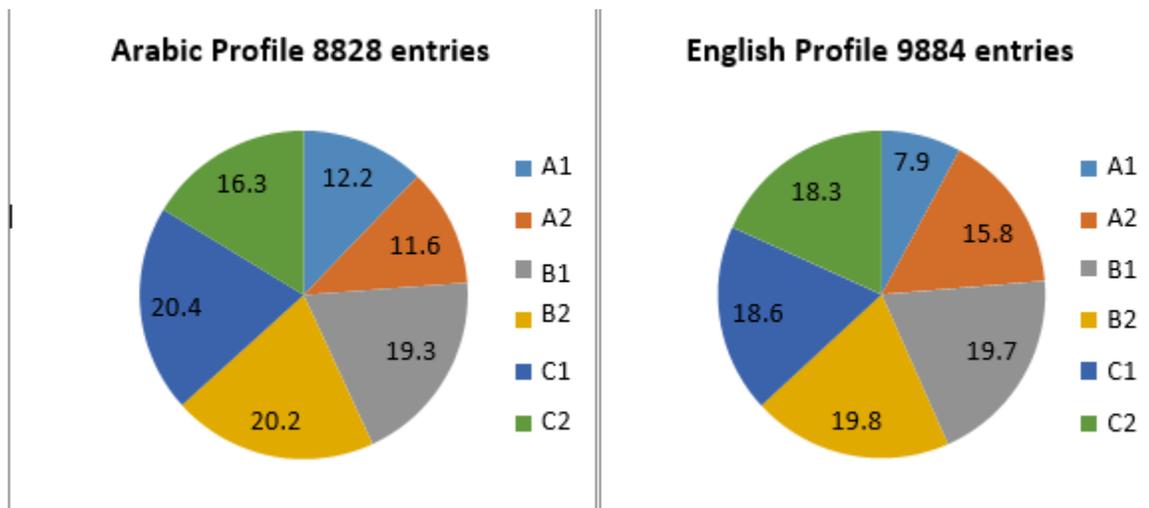
Vocabulary lists are essential resources in various language studies, ranging from language learning to all applied linguistics disciplines. Corpus-derived frequency lists are the common way of producing vocabulary lists, in which “words are arranged according to the number of times they occur in particular samples of language” (Richards, 1974). Sharoff et al. (2014) stated that, a key pedagogic problem with using frequency lists is that they differ according to the corpus source they were extracted from. If a corpus contains specialist texts, some technical words are listed at higher positions in the lists at the expense of everyday basic words. They argued that having a pedagogical reference for second language learning can isolate such special words.

The aim of this research was to compile a new Arabic vocabulary list from available Arabic word lists and to classify the newly developed list with the Common European Framework of Reference for Languages proficiency (CEFR). CEFR is an international standard for evaluating the language ability and has been widely used for the curriculum design and assessments for many European and non-European languages (Soliman, 2018).

In this research we merge two Arabic word lists derived from corpora: the Buckwalter and Parkinson Arabic frequency list (Buckwalter and Parkinson, 2014), and the Arabic Kelly’s list (Kilgarriff et al., 2014). We add 4024 words extracted from the lists presented at the beginning of each chapter in ‘Al-Kitaab’ (Brustad et al., 2013) a text book for teaching Arabic as a second language considering it as the pedagogical reference. On one hand, Buckwalter and Parkinson vocabulary list was introduced in a book entitled “A frequency dictionary of Arabic: Core vocabulary for learners” with 5000- Arabic word frequency based on 30-millionword corpus of academic/nonacademic and written/spoken texts. On the other hand, Arabic KELLY’s list is 9,000 produced from the Kelly project in Leeds (Kilgarriff et al., 2014) as a part from nine languages bilingual vocabulary lists the Arabic list, it was obtained from a 170million-word internet-based corpus and it is a frequency word list represented in lemmas associated with their CEFR levels.

Hence, the aim of this research was to compile an Arabic list containing only MSA variety. We considered only MSA words from Buckwalter and Kelly’s list. A comparison between these two lists shows inconsistency between the lemmas entries which make it difficult to align them. Though, to validate the results we analysed the words presented in all lists by MADAMIRA, a robust Arabic morphological analyser and part of speech tagger (Pasha et al., 2014). The results show that some words which appear in the top 1000 frequent words in Buckwalter’s list are classified at the superior level in Kelly’s list such as ‘صهيوني’ ‘Zionist’. These results urge us to rely on the Kelly’s list classification for the words that does not exist in ‘Al-Kitaab’ list. The compilation of the final list started with the 4024 words from ‘Al-Kitaab’ book chapters. Then we averaged 3669 intersected lemmas in both Kelly’s and Buckwalter lists. After that manually classify 525, 616 lemmas from Kelly’s and Buckwalter lists respectively. This resulted in a new classified Arabic Vocabulary list consisting of 8834 distinct lemmas. When compared with the

English profile word list, the Arabic list demonstrated that words are approximately equally distributed across the CEFR levels as shown in figure 1.



Evaluation of Association Measures in Automatic Extraction of Collocations in Arabic

Souhaila Messaoudi (University of Leeds), James Dickins (University of Leeds) & Claire Brierley (University of Leeds)

An association measure is a mathematical formula which identifies candidates forming reliable collocations (Pecina 2010), (Gries 2013). The aim of this study is twofold: first, to measure the efficiency of Log Dice, a relatively new measure to automatic extraction of collocations in general and in Arabic in particular; second, to highlight the relationship between collocational dimensions and four different association measures in corpus linguistics to provide insights on their practical effects. The association measures used in this experiment are: T-score, MI, Log Likelihood, and Log Dice, with an emphasis on the last two. Moreover, this study classifies the examined association measures in accordance with their approach of extracting collocations as: predominant frequency approach measures and predominant exclusivity approach measures. In this regard, T-score and Log Likelihood, which share the frequency approach, are categorized as predominant frequency approach measures, while MI and Log Dice, which adopt the exclusivity approach, are the predominant exclusivity approach measures.

The experiment first reports the scores of the four association measures in extracting the top ten ‘adjective plus noun’ collocations from Arabic Web Corpus (arTenTen), one of the biggest corpora in Modern standard Arabic (MSA). It then compares the scores of Log Likelihood, which received wide appreciation in extracting collocations in general and in Arabic in particular (Evert, 2008; Boulaknadel et al., 2008; Thanopoulos, Fakotakis, and Kokkinakis, 2002), to those of Log Dice, which has not been used for automatic extraction of Arabic collocations as seen in the previous literature.

Using Sketch Engine, a chosen list of focal words representing Arabic equivalent nouns obtained from a previously compiled list of friends between English and French (Thody, Evans, and Rees, 1985) was explored. For each of the words in the list, the best adjectival collocates were found. These target collocates, which included technical terms, function words and punctuation in this analysis, were considered only if they occurred in the top ten collocates of the Log Likelihood and Log Dice score lists respectively. Moreover, an adjectival collocate was to be considered a perfect collocate if it occurred in both lists with the same rank. Findings of this study show that Log Dice scores do not extract as many technical terms as the MI scores do. This asserts that in theory and practice Log Dice tries to compensate for the bias of low frequency which is the main side effect of MI (Gablasova, Brezina, and McEnery, 2017) by highlighting exclusive, though not necessarily rare, collocates. Therefore, Log Dice can be regarded as the best association measure for exclusivity as it outperforms MI. Moreover, Log Likelihood and T-score seemingly gave very similar results for the best scoring collocates (e.g., they both render punctuation and function words). However, Log Likelihood does not highlight free word combinations as the T-score does. Finally, this study demonstrates the effectiveness of using Log Dice in the extraction of Arabic collocations. Although Log Likelihood is said to be the best measure among all the associations of frequency, the precision of Log Dice outperforms that of Log Likelihood as the former performs effectively even on raw (not pre-processed corpora).

Facing Time: A study of the spatial representation of time in Classical vs. Modern Arabic

Mariam Aboelezz (Lancaster University)

Time is an abstract concept which is represented differently in different languages. Most languages, like English, make use of horizontal spatial metaphors where time moves forward. Studies have shown that writing direction influences the direction in which time is imagined to move in a given language: in English, which is written from left to right, time also moves from left to right, while Arabic, a right-to-left language, represents time from right to left. Nevertheless, the trajectory is the same: in both languages, as in most of the world's languages, we move forward into the future, 'facing' it. The English expression "I look forward to meeting you tomorrow" is readily translatable into Arabic without the need to adjust the spatial metaphor.

Arabic, at least today then, is a 'future-facing' language. In this paper I argue that this was not always the case. As a starting point, I analyse spatial representations of time in the Quran to demonstrate that these metaphors clearly and consistently convey a 'past-facing' trajectory. I then turn to the question of how/when the transition in the trajectory of time occurred in Arabic. Focusing on the words employed to encode time in the Quranic analysis, I trace their dictionary definitions through Arabic sources from *kitāb al-ʿayn* (8th C) to *muḥīṭ al-muḥīṭ* (19th C). The examples from pre-Islamic and early Islamic Arabic poetry provided in the earlier dictionaries lend credence to the hypothesis that the Arabic representation of time was past-facing. By contrast, later *nahḍah*-era dictionaries introduce entries such as *mustaqbal* (future; literally: that which is faced) which imply that the transition in trajectory has occurred. I reflect on the implications of these findings for the translation of the Quranic text, and of Classical Arabic texts more broadly, and argue that they provide compelling support for what is occasionally deemed an 'artificial' distinction between Classical Arabic and Modern Standard Arabic.

Needs Analysis and Course Design in Arabic for Academic Purposes: A Case Study at a Saudi Arabian University in Riyadh

Soha Altayar (University of Edinburgh)

The needs analysis (NA) is a fundamental tool in language course design because it helps determine students' needs in the target learning contexts. However, in the field of teaching Arabic for Academic Purposes (AAP), most language courses are designed to rely on the teachers' experiences and intuition, with no attention given to students' individual needs. This may explain the shortage of Arabic linguistic studies that consider NA and course evaluation. Thus, by using the case study methodology and the learning-centred approach (LCA), the present study aimed to analyse the academic targets (what is lacking, necessities and wants) and learning needs of overseas students studying at a female university in Riyadh, Saudi Arabia. To this end, a triangulation method of data collection was used, including semi-structured interviews with L2 Arabic undergraduate students in different academic fields, focus group interviews with AAP teachers and official AAP documents, which were collected to gain further insight. The findings of the analysis showed that both students and teachers agreed that listening to and comprehending lectures are the most needed skills for students. Similarly, both groups indicated that students lack competence in planning and writing academic research before they move to the university level. The results also showed that students want to improve their verbal communication skills to interact effectively with native speaking colleagues, both inside and outside classrooms. The analysis of the learning needs found that religious needs are essential, and they motivate students to learn Arabic as an L2. It was also shown that Arabic for General Academic Purposes courses are more likely to fulfil students' target academic needs than are Arabic for Specific Academic Purposes courses.

Social Media language resources for Arabic dialect research: an evaluation

Eric Atwell (University of Leeds)

Arabic dialectology has traditionally focused on listening to informants and audio recordings of dialect speakers as sources of evidence. Arabic dialect speakers are taught formal writing in Modern Standard Arabic, hence until recently little dialect text existed. Social media like Twitter and FaceBook have enabled the general public to write informally including in dialects, and this offers a new data source for Arabic dialectology. Advantages include:

- it is relatively easy and cheap to collect large amounts of unscripted spontaneous informal dialect data: millions of words of text can be scraped from social media such as Twitter and FaceBook [1,2].
- there is no need to transcribe, the data is already text
- we can collect comparable text corpora for a range of dialects, to enable comparative analysis.
- Research on dialects in social media is attracting interest from Computing academics and industry, bringing new resources and funding to Arabic dialectology.
- Machine Learning and Data Mining methods can be directly applied to large dialect text corpora from social media, to automate the learning of linguistic patterns and insights. For example, there is growing research in automated dialect classification software [3,4].

But there are also disadvantages:

- some dialectology research focuses on phonetic variation; this is captured only indirectly in dialect text.
- Due to lack of conventions for writing Arabic dialects, there can be variations in individual transcription/spelling.
- It is challenging to identify a set of tweets representative of a specific dialect. One strategy is to compile a “seed-list” of words used only in the given dialect, then collect only tweets including one or more of these dialect words; but this leaves out many tweets by speakers of the given dialect, and results in a corpus with word frequencies skewed towards the seed words. Another strategy is to collect all tweets from specific locations, for example the main cities in a dialect region; this can include dialect tweets lacking the seed terms. However this also includes tweets by visitors who are not dialect speakers (for example not all tweets in Riyadh are Najdi Arabic speakers); and it leaves out dialect tweets from outside main population centres, and the diaspora of dialect speakers who have left home.
- Tweets may include non-dialect data, such as code-switching between dialect, MSA, and/or English/French/etc; emojis, URLs and other nonlanguage. It is particularly difficult to confidently identify code-switching between MSA and dialect in written text.

We conclude there is room in Arabic dialectology research for both traditional approaches based on audio recordings, and new approaches based on social media dialect text analytics.

Cross-cultural Pragmatics: apology Strategies in Libyan Arabic

Ahmed Elgadri (Lancaster University) & Bethan Davies (University of Leeds)

In the last twenty years, studies on cross-cultural pragmatics in general and apology strategies in specific have focused on western and East-Asian pragmatics. A small volume of research has been conducted in investigating speech acts production by Arabic dialect speakers. Therefore, this study investigated apology strategies used by Libyan Arabic speakers using an online Discourse Completion Task (DCT) questionnaire. The DCT consisted of six situations covering different social contexts. The survey was written in Libyan Arabic dialect to help generate vernacular speech as much as possible. The participants were 25 Libyan nationals, 12 females, and 13 males. Also, to get a deeper understanding of the motivation behind the use of certain strategies, the researcher interviewed four participants using Libyan Arabic dialect as well. The results revealed a high use of IFID, offer of repair, and explanation. Although this might support the universality of apology strategies, it was clear that cultural norms and religion determined the choice of apology strategies significantly. This led to the discovery of new culture specific strategies as outlined in the paper. This shows the importance of culture. This research gives an insight into politeness strategies in the Libyan community and it is hoped to contribute to the field of cross-cultural pragmatics.