

# Table of Contents

<b>ACKNOWLEDGEMENTS AND WELCOME .....</b>	<b>2</b>
<b>ABOUT THE ALS .....</b>	<b>3</b>
<b>PROGRAM .....</b>	<b>4</b>
<b>ABSTRACTS</b>	
<b>THURSDAY SESSIONS .....</b>	<b>8</b>
<b>FRIDAY SESSIONS .....</b>	<b>27</b>
<b>SATURDAY SESSIONS .....</b>	<b>49</b>

## THE TWENTY-SEVENTH ANNUAL SYMPOSIUM ON ARABIC LINGUISTICS

Indiana University  
Bloomington, Indiana

February 28 – March 2, 2013

### Acknowledgements

Host Institutions:

**INDIANA UNIVERSITY**  
DEPARTMENT OF LINGUISTICS  
CENTER FOR THE STUDY OF THE MIDDLE EAST  
&  
The ARABIC LINGUISTICS SOCIETY

Dear Attendees

Welcome to Bloomington and Indiana University. We are pleased to host the 27th meeting of the Arabic Linguistics Symposium, the most prestigious conference on Arabic Linguistics. This is the first time that the Symposium is being held in Bloomington and we hope that you make most of your stay. The University is known for the wide number of languages that are taught on campus and its many outstanding language departments. In making this event possible, we want to acknowledge the support of the Center for the Study of the Middle East and the Department of Linguistics at Indiana University, the two co-sponsors. Both have allocated financial resources to make the Symposium possible. In particular we want to thank Liese Hilgeman, Myriem Benzouina, Michael Hancock-Parmer and Çiğdem Balım of the Center and Professor Robert Botne of the Department of Linguistics. In addition we thank the IU Conference Services for help with the registration and John Benjamins Publishing Co. for their support of publishing the conference volumes. Most importantly, we thank the executive board of the Arabic Linguistics Society and especially its executive director Mushira Eid for reviewing abstracts and their help with organizing all aspects of the conference. Finally, we would like to thank you all for choosing to share your research and to learn from others at ALS 27. We look forward to what promises to be an engaging and rewarding Symposium.

Sincerely,

Feisal Istrabadi and Stuart Davis

## About the Arabic Linguistics Society

The Arabic Linguistics Society (ALS) was founded in June 1988 as a non-profit organization for the purpose of encouraging research and the sharing of research in the field of modern Arabic linguistics. The Society provides a forum for scholars interested in the study of Arabic within current linguistic theories and analyses. As such, it is the major, if not only, professional society in North America that is exclusively dedicated to the advancement of research on Arabic linguistics and plays a critical role in supporting and disseminating linguistic scholarship on Arabic.

The Society sponsors an annual symposium on Arabic linguistics, the first held in 1987 hosted by the University of Utah, Salt Lake City. Since then over 20 other major universities have hosted the annual symposium. In 1998, the Society sponsored its first International Symposium on Arabic Linguistics to maintain contact with Arabic linguists at universities abroad and encourage international research on Arabic linguistics. It continues to sponsor international symposia on a less regular basis.

For disseminating research on Arabic linguistics, the Society sponsors the publication of papers selected from those presented at its symposia in a series entitled *Perspectives on Arabic Linguistics* and published by John Benjamins (Amsterdam and Philadelphia). Papers are peer-reviewed and edited prior to final acceptance for publication.

### Executive Board (2010-2013)

Reem Khamis-Dakwar, Adelphi University  
Mustafa Mughazy, University of Western Michigan  
Hamid Ouali, University of Wisconsin-Milwaukee  
Dilworth Parkinson, Brigham Young University  
Usama Soltan, Middlebury College

Executive Director:  
Mushira Eid, University of Utah

## THURSDAY, FEBRUARY 28, 2013

- 7:30-8:15      **Registration**  
8:15            **Welcome/Opening Remarks**

### Phonology I (Chair: Kenneth de Jong, IU)

- 8:30-9:00      **Vocalic Length in One Semitic Language: The case of Egyptian Arabic vowel system**  
*Radwa Fathi, Université Paris 7 /LLF- CNRS*
- 9:00-9:30      **An Acoustic Study of Epenthetic Vowels in Lebanese Arabic**  
*Nancy Hall, California State University, Long Beach*
- 9:30-10:00     **The Acoustic Correlates of Emphatic Geminates in Jordanian Arabic**  
*Mohammad Al-Masri, University of Oklahoma*
- 10:00-10:30    **Acoustic Correlates and Perception Cues of Primary Emphatics in Beirut Arabic**  
*Emilie Durand-Zuniga, University of Texas at Austin*
- 10:30-10:45    **Coffee Break**

### Phonology II (Chair: Kenneth de Jong, IU)

- 10:45-11:15    **The Status of *Daad* in a Southwest Arabian dialect (Tihami Qahtani)**  
*Khairia Al-Qahtani, University of Essex, UK*
- 11:15-11:45    **To Metathesize or Not to Metathesize: Sonority Constraints in Tunisian Arabic Nouns**  
*Suyeon Yun, Massachusetts Institute of Technology*
- 11:45-12:45    **Keynote Address:  
Explorations at the syntax-phonology interface in Arabic**  
*Sam Hellmuth, University of York, UK*
- 12:45-1:45     **Lunch**

### Syntax I (Chair: Youssef Haddad, University of Florida)

- 1:45-2:15      **On the Featural Properties of Complementizers in Arabic**  
*Hamid Ouali, University of Wisconsin-Milwaukee*
- 2:15-2:45      **On the Syntax of Spatial 'Prepositions' in Lebanese Arabic**  
*Lina Choueiri, American University in Beirut*
- 2:45-3:15      **On the Syntax of *ʔillaa* in Egyptian Arabic**  
*Usama Soltan, Middlebury College*
- 3:15-3:30      **Coffee Break**

## Semantics (Chair: Mushira Eid, University of Utah)

- 3:30-4:00      **The Semantic Complexity of Kuwaiti Arabic: The case of *dašš***  
*Yousuf AlBader, University of Sheffield*
- 4:00-4:30      **Identifying Semantic Relations in Arabic**  
*Sameh Alansary, Bibliotheca Alexandria*
- 4:30-4:45      **Coffee Break**
- 4:45-5:45      **Keynote Address:**  
**Arabic Verbal And Nominal Plurals And The Syntax Morphology Interface**  
*Elabbas Benmamoun, University of Illinois, Urbana-Champaign*
- 6:00-7:30      **Reception, Indiana University Art Museum**

## FRIDAY, MARCH 1, 2013

## Syntax II (Chair: Usama Soltan, Middlebury College)

- 8:30-9:00      **On the Clausal Anchoring of Parenthetical Verb Phrases in Moroccan Arabic**  
*Noureddine Elouazizi, Simon Fraser University*
- 9:00-9:30      **Attitude Datives in Lebanese Arabic: Pronouns that merge too high to be bound**  
*Youssef Haddad, University of Florida*
- 9:30-10:00     **A Minimalist Approach to Restrictive and Free Relatives in MSA**  
*Abdulrahman Alqurashi, University of Essex*
- 10:00-10:15    **Coffee Break**

## Discourse/Corpus Analysis (Chair: Reem Khamis-Dakwar, Delphi University)

- 10:15-10:45    **A Saliency-based Analysis of the Tunisian Arabic Demonstrative *hāk* as Used in Oral Narratives**  
*Amel Khalfaoui, Florida Atlantic University*
- 10:45-11:15    **Converging Linguistic Evidence: The synonymy of Arabic COME verbs**  
*Dana Abdulrahim and Antti Arppe, University of Alberta*
- 11:15-12:15    **Keynote Address:**  
**Modeling Sociopragmatic Language Use in Social Media in Arabic and English: A Comparative Computational Perspective**  
*Mona Diab, George Washington University*
- 12:15-1:30     **Lunch**  
**Arabic Linguistics Society Business Meeting**

## Sociolinguistics & Variation (Chair: Clancy Clements, IU)

- 1:30-2:00      **Computational Measures of Linguistic Variation**  
*Mahmoud Abunasser and Elabbas Benmamoun, University of Illinois at Urbana-Champaign*
- 2:00-2:30      **Glottal-initial Verbs 'axad/'akal in Contact Situations: Data from Amman Arabic**  
*Enam Al-Wer, University of Essex, and Hanadi Ismail*
- 2:30-3:00      **Urbanization and Linguistic Change in Jeddah**  
*Aziza Al-Essa, King Abdulaziz University*
- 3:00-3:30      **French Nouns in CS with Two Arabic dialects: A comparison of community norms**  
*Rebekah Post, University of Texas at Austin*
- 3:30-3:45      **Coffee Break**
- 3:45-4:45      **Keynote Address:**  
**The Maghreb-Mashreq language ideology and the politics of identity in a globalized Arab world**  
*Atiqah Hachimi, University of Toronto, Scarborough*
- 4:45-5:00      **Coffee Break**

## Historical Linguistics (Chair: Hamid Ouali, University of Wisconsin, Milwaukee)

- 5:00-5:30      **The Antecedents of Arabic Dialects: A speech communities approach**  
*Alexander Magidow, University of Texas at Austin*
- 5:30-6:00      **Grammaticalization of the Motion Verb 'rah' as a Prospective Aspect Marker in Syrian Arabic**  
*Najib Jarad, University of Sharjah*
- 6:00-6:30      **Jiim and the Class of Sun Letters: A Historical and Dialectological Perspective**  
*Aaron Freeman, University of Pennsylvania*
- 6:30-7:00      **The Evolution of *Dād***  
*Salman Al-Ani and Anthony Woodhams, Indiana University*

## SATURDAY, MARCH 2, 2013

### Acquisition (Chair: Dil Parkinson, Brigham Young University)

- 8:30-9:00      **Compensatory Lengthening: Evidence from Child Arabic**  
*Eman Abdoh, King Abdulaziz University*
- 9:00-9:30      **Linguistic Transfer in Learning English as a Second Language in Typically Developing Arabic Heritage School-Age Speakers**  
*Reem Khamis-Dakwar, Adelphi University, and Elabbas Benmamoun, University of Illinois at Urbana-Champaign*
- 9:30-10:00     **Can L2 Primes Trigger L1 Translation Targets in Mased Priming?: New Evidence from a highly proficient English-Arabic Bilingual**  
*Mahmoud Azaz and Kenneth Forster, University of Arizona*
- 10:00-10:30    **Gender and Plausibility in the Disambiguation of Relative Clauses in L2 Arabic**  
*Abdelaadim Bidaoui, Rebecca Foote, and Mahmoud Abu Nasser, University of Illinois at Urbana-Champaign*
- 10:30-10:45    **Coffee Break**

### Experimental and Interlanguage Studies (Chair: Çiğdem Balım, IU)

- 10:45-11:15    **Verb Inflections in Adolescents with Down Syndrome: Experimental approach**  
*Khawla Aljenaie, Kuwait University*
- 11:15-11:45    **Agrammatism in Moroccan Arabic-Speaking Subjects**  
*Samir Diouny, Chouaib Doukkali University, Morocco*
- 11:45-12:15    **An Exploration of the Formation of Canonical Agreement Morphosyntactic Features in English-Arabic Interlanguage**  
*Boshra El-Ghazoly, Indiana University*

### Special Session:

#### Arabic in Computer-Mediated Communication (Chair: Stuart Davis, IU)

- 12:45-1:15     **Social Media Arabic**  
*Muhammad Abdul Mageed, Indiana University*
- 1:15-1:45      **Variation in the Representation of Arabic Consonants in Facebook**  
*Duaa Abu Elhija Mahajna, Indiana University*
- 1:45-2:15      **Arabic Chat Alphabet: A data-oriented analysis of variation in Latinized Arabic**  
*Paul Rodrigues, C. Anton Rytting, and Timothy Buckwalter, University of Maryland*

## Session: Thursday, 8:30 AM

### Vocalic Length in One Semitic Language: The case of Egyptian Arabic vowel system Radwa Fathi, Université Paris 7 /LLF- CNRS

Egyptian Arabic (henceforth EA) displays an array of surface vocalic qualities including both short and long counterparts: [i], [i:], [u], [u:], [a], [a:], [e], [e:], [o], [o:] amongst other variants resulting from the vicinity of emphatic consonants and gutturals. Most studies concerned with the makeup of the vowel system of EA<sup>1</sup> assume vowel length to play a contrastive role, hence an underlying system such as follows: /i/, /i:/, /u/, /u:/, /a/, /a:/, /e:/, /o:/.

- (1)
- |         |          |
|---------|----------|
| i /i:   | u /u:    |
| e: (<a) | o: (<aw) |
| a /a:   |          |

This underlying vowel inventory could be outlined as follows under the assumptions of the Theory of Elements (KLV, 1985):

- (2)
- |                |   |   |   |   |   |
|----------------|---|---|---|---|---|
| Round/Back     | I | U | v | I | U |
| High           | v | A | A | A | A |
| Segmental tier | i | u | a | e | o |

The distributional analysis of EA vowels shows an unusual asymmetry. While short vowels are freely distributed in the word, they can occur in ultimate, penultimate and antepenultimate positions; long vowels by contrast seem to be relatively restricted, they can be found in either the ultimate or the penultimate positions. This system poses a number of problems.

First, the inclusion of long mid-vowels and the exclusion of short mid-vowels raise the following questions:

1. Why does EA generate three short vowel qualities /i/, /u/, /a/ when it can effectively generate five long ones /i:/, /u:/, /a:/, /e:/, /o:/?
2. Why can length contrast be established for peripheral vowels and not for mid-vowels?
3. Is vowel length contrast quality-dependent?
4. Is vowel length contrast getting lost in the evolution of the EA vowel system?
5. Universal tendencies point to the higher expectancy of phonemes with less marked features to occur, and to length in vowels to exhibit the characteristics of a marked feature. Why does EA contravene these universals and exhibits the marked phonemes /e:/ and /o:/ in the absence of the less marked ones /e/ and /o/?

Second, in a language in which vowel length is supposedly contrastive, the distribution of long vowels is presumably free as well as the positions where length contrast can be possibly established. What about this restricted distribution of both: 1) long vowels and 2) length contrast?

In my treatment of these issues, I review previous claims on the underlying vowel inventory in order to show that these claims do not adequately characterize the phonological status of both: 1) mid-vowels, and 2) vocalic length. Following the non-linear model CVCV (Lowenstamm, 1996) and the Theory of Internal Structure of Phonological Elements (KLV, 1985) I present an argument that a long mid-vowel effectively stems from the same segmental materials underlying a diphthong (as in 3), the following hypothesis is thus formulated: long mid-vowels in EA are not phonemes on their own; they are but the surface reflection of two underlying phonemes.

- (3)
- |       |     |       |      |
|-------|-----|-------|------|
| V C V | V C | V C V | V C  |
|       |     |       |      |
| A.I   | A I | A.U   | A U  |
| [e:]  | [a] | [o:]  | [aw] |

<sup>1</sup> Studies include: Harrell (1957) Broselow (1976), Gadalla (2000) and Watson (2002) amongst others.

With mid-vowels excluded from the list of phonemes, a three-vowel system can finally be posited for EA:

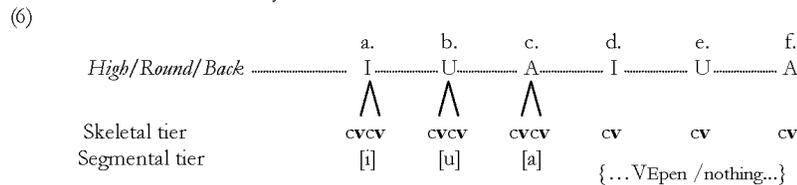


I then revisit vowel length contrast and I show that the bases upon which it is often established are far from being reliable. What I claim is that vocalic length in this language is a function of templatic space. Consider the piece of data in (5). Where a word is present in both classical Arabic (henceforth CA) and EA, EA retains long vowels of CA and doesn't realize short ones. They either disappear or get replaced by an epenthetic vowel [e].

(5)

	a. Classical Arabic	b. Egyptian Arabic	Gloss
i.	labisu:	lebsu	'they put on'
ii.	ʔuktubi:	ʔektebi	'write! Imperative (2sg.f.)'
iii.	ka:bu:s	kabu:s	'nightmare'
iv.	taktubu	tekteb	'you (sg.m.) are writing down'
v.	laʕib	leʕb	'playing (verbal noun)'

In view of that, vowel length is redefined so that it is no more the vowel that possesses length or shortness per se; rather I show that length is a function of templatic space. Peripheral vowels must branch onto two vocalic positions otherwise they delink. I argue and I support by evidence that the EA vowel system can be characterized as follows:



This hypothesis is supported in EA by the case of allomorphy of the singular feminine marker /at/ which alternates between [a]: ga:ra 'neighbor (sg.f.)', and [(e)t]: garti 'my neighbor (sg.f.)' and gartna 'our neighbor (sg.f.)'. It will be shown that it is either the vocalic component that surfaces when the peripheral vowel *a* finds the space to branch (as in 6.c), otherwise it dissociates under the linking of the consonantal component *t* (as in 6.f).

### References

Broselow, E. (1976). The Phonology of Egyptian Arabic. Doctoral dissertation. University of Massachusetts, Amherst.

Gadalla, H. (2000). Comparative Morphology of Standard and Egyptian Arabic. Muenchen: Lincom Europa.

Harrell, R. (1957). The phonology of Colloquial Egyptian Arabic. New York: American Council of Learned Societies.

Kaye, J., Lowenstamm, J. and Vergnaud, J.R. (1985). The Internal Structure of Phonological Representations: A Theory of Charm and Government. *Phonology Yearbook*, 2, 305-328.

Lowenstamm, J. (1996). CV as the Only Syllable Type, *Current Trends in Phonology, Models and Methods*. J. Durant and B. Laks (eds.), Salford, Manchester, ESR.

Watson, J. (2002). The Phonology and Morphology of Arabic. Oxford: Oxford University Press.

**Session: Thursday, 9:00 AM**

**An Acoustic Study of Epenthetic Vowels in Lebanese Arabic**

*Nancy Hall, California State University, Long Beach*

An acoustic study of lexical vowels in Lebanese Arabic Levantine Arabic speakers often insert epenthetic [i] in final CC clusters, as in /bint/ [binit] 'girl'. Traditionally, these vowels are assumed to be phonetically indistinguishable from lexical [i]. However, this acoustic study finds variation in epenthetic vowel quality, with epenthetic vowels tending to be low and centralized compared to supposedly homophonous lexical [i].

This finding has implications for theoretical phonology. Levantine is typically described as having opaque interactions between epenthesis and other processes (that is, epenthesis alters the contexts that guided processes such as stress, syncope, and vowel shortening), but if epenthetic vowels are phonetically distinguishable from lexical vowels, then the interactions are not truly opaque.

I recorded 22 Lebanese speakers producing 8 repetitions of 15 minimal or near-minimal pairs, such as the verb [libis] 'he wore' (where both vowels are lexical) versus the noun [libis] 'clothes' (where the second vowel is epenthetic). This study used a novel method of elicitation, in which participants read the target word after hearing a colloquial sentence containing the word (with the target word itself replaced with white noise, to avoid providing a model for pronunciation). This method was successful in eliciting colloquial pronunciations, which is often a problem in laboratory work on Arabic phonetics.

The formant values and durations of the second-syllable vowels were compared using Praat. The results show variation among speakers: about a third produce an epenthetic vowel that is acoustically identical with lexical [i], another third produce a schwa-like epenthetic vowel that is categorically distinct from lexical [i], and the rest produce clouds of epenthetic and lexical vowel tokens that are significantly different, yet strongly overlap. There is no discernible demographic pattern as to which speakers fall into which group. Nor is there any correlation between how frequently a speaker epenthesizes and whether they differentiate epenthetic vowels.

The most common direction of difference is higher F1 and lower F2 in epenthetic vowels (i.e., the epenthetic vowel is lower and more centralized than the lexical vowel), with the F2 difference more robust. However, a few speakers show an F1 difference in the opposite direction, and those speakers do not have a significant F2 difference. Durational differences were (surprisingly) not significant. First-syllable vowels also show a difference in quality, but this is weaker and likely due to co-articulation.

This study has implications for phonological analyses of vowel epenthesis. If acoustic cues exist to indicate which vowels are epenthetic and which are lexical,

then interactions traditionally described as ‘opaque’ actually are not: it is possible to accurately reconstruct the underlying form based on the surface form alone. This finding may shed light on how speakers learn the underlying representations and the complex phonological mappings to surface forms.

I will also present preliminary results from a similar study of 120 speakers of Palestinian Arabic, which was conducted in Tel Aviv and Haifa in summer 2013. Due to its larger size, this study may be able to shed more light on the demographic factors that affect epenthetic vowel quality. Besides comparing epenthetic and lexical vowels, this study compares underlyingly short vowels to vowels that are shortened by phonological rule (for example, [dZabes], from /dZaab-S/ ‘he did not bring’, versus [kabeS], from /kabS/ ‘ram’). I hypothesize that here, too, there may be an incomplete neutralization of underlying distinctions.

**Session: Thursday, 9:30 AM**

**The Acoustic Correlates of Emphatic Geminates in Jordanian Arabic**  
*Mohammad Al-Masri, University of Oklahoma*

This paper investigates whether emphatic geminates show different effects of emphasis compared to the already established acoustic characteristics of emphasis triggered by single consonants. Previous studies have established several of these characteristics, most noticeably, a significant lowering of F2 and a rise of F1 and F3 in vowels adjacent to emphatic consonants, (Card: 1983; El-Halees: 1985 & 1987; Herzallah: 1990; Al-Masri and Jongman: 2004; Jongman, et.al. 2007 & 2011 among others). Characteristics of emphasis have not been established for geminates. Arabic generates different phonemes based on quantitative, not only qualitative differences, (Flege and Port: 1981; Munro: 1993; Hassan 2002; De Jong and Zawaydeh: 2002; De Jong: 2004). Consider the following examples, \* (dots are used as syllable breaks; underlined consonants are emphatic):

1. /ka.sar/ he broke                    /kas.sar/ he smashed
2. /ba.tal/ hero                        /bat.tal/ he quit
3. /ba.far/ humankind                /baf.far / he passed good news

32 pairs of CV.CVC words and non-words with the geminate target consonant in word-medial position have been read in a carrier sentence by 6 native speakers of Jordanian Arabic (3 males and 3 females). Since previous studies show that a change in vowel frequency is the most salient effect of emphasis, only vowel frequency data (F1, F2, and F3) were analyzed. Vowel frequencies were measured at three points (vowel onset, vowel midpoint and vowel offset). Results were averaged across three repetitions for each speaker.

Preliminary results suggest that geminates trigger greater effects of emphasis, mostly lowering of F2 value, as compared to single consonants. These results are equally true for short and long vowels – the latter being reported in some studies to block effects of emphasis, though the magnitude of F2 lowering was significantly greater in the environment of short vowels preceding the target consonant. F1 and F3 rise significantly especially when measurements are taken closer to the target consonant. As reported in several previous studies on emphasis in single consonants, no clear effect of gender was observed.

Results of the present study provide new evidence for the acoustic behavior of a set of Arabic consonants. Further, results provide educators with research-based findings that can help better highlight the distinctions between these sounds utilizing vowel information, which minimizes the resources of one of the major difficulties arising from the learning of Arabic speech sounds.

**Session: Thursday, 10:00 AM**

**Acoustic Correlates and Perception Cues of Primary Emphatics in Beirut Arabic**  
*Emilie Durand-Zuniga, University of Texas at Austin*

Acoustic Correlates and Perception Cues of Primary Emphatics in Beirut Arabic Emphasis is a feature of Arabic that has received a fair amount of attention in phonology studies for several decades now. From an articulatory perspective, it is characterized by a secondary constriction in the posterior vocal tract with a primary constriction usually in the front of the mouth (dental/alveolar region) (Davis, 1991). Arabic has five primary emphatics, namely /t<sup>ʕ</sup>, d<sup>ʕ</sup>, s<sup>ʕ</sup>, z<sup>ʕ</sup>, ʃ<sup>ʕ</sup>/, described as independent phonemes which form minimal pairs with their non-emphatic counterparts (Younes, 1994), though no one variety of Arabic uses all five (Davis, 1991). Acoustically, the main characteristic of emphasis in the dialects studied so far is a lower F2 value in the part of the vowel that is closest to an emphatic consonant as opposed to a plain consonant (Khattab et al., 2006; Jongman et al., 2011). However, F1 and F3 offer a much less homogeneous picture than F2: some have found a higher F1 and F3 in vowels neighboring emphatic consonants as opposed to plain consonants (Jongman et al., 2011), while others reported no consistent influence of emphasis on F1 and F3 (Card, 1983).

The object of this pilot study is to identify the role of vowel formants in both the production and perception of emphasis in Beirut Arabic, a dialect that has so far not been investigated. Two native male Beirutis participated in this research, one for the production experiment and one for the perception experiment. In the production study, the four emphatics of Beirut Arabic /t<sup>ʕ</sup>, d<sup>ʕ</sup>, s<sup>ʕ</sup>, z<sup>ʕ</sup>/ and their plain counterparts /t, d, s, z/ were recorded in target words that contained the target consonant in word-initial position followed by one of three vowels (a, i, u). The stimuli were audio recordings of one word each, pronounced in Modern Standard Arabic, and the participant was instructed to produce each of those tokens in his native Beirut Arabic in a carrier sentence. In each of the 240 tokens recorded, the first three formants of the vowel following the target consonant were measured at the first regular vocal pulse of F2 in the vowel. The ANOVA analysis indicated that F2 was lower and F1 higher after emphatics, but it did not indicate a main effect on F3.

In the perception study, the participant heard 250 tokens of /ti:/ and had to indicate where each token contained an emphatic consonant, by clicking one of two boxes on a computer screen: the Arabic letter for emphatic [t<sup>ʕ</sup>] and the Arabic letter for plain [t]. F1 and F2 values were manipulated in the first 150ms of the vowel, starting at the first regular vocal pulse in F2. F1 and F2 were given five different values and each F1 value was matched with each F2 value, resulting in 25 individual tokens, which were played 10 times each in randomized order. The logistic regression analysis indicated that F2 was a stronger acoustic cue than F1 in the perception of emphatics in Beirut Arabic.

**Session: Thursday, 10:45 AM**

**The Status of *Daad* in a Southwest Arabian dialect (Tihami Qahtani)**

*Khairia Al-Qahtani, University of Essex, UK*

A widely accepted assumption in Arabic dialectology is that none of the modern Arabic dialects maintain a phonemic distinction between Daad and Dhaa الضاد والظاء (see Al-Wer 2004). Equally, the Old Arabic and proto-Semitic pronunciation of Daad, which according to Sibawayhi was an emphatic voiced lateral fricative, is considered obsolete (see Steiner 1976; Corriente 1978).

In recent years, there have been reports (see for example Al-Azraqi 2010) that a lateral pronunciation of Arabic Daad can still be found in southern Arabian dialects, especially in isolated regions. Nonetheless, the phonemic status of this sound vis a vis Dhaa is unclear, i.e. whether the two sounds are actually phonemically distinctive. In this paper, I will present data from my research in Tihami Qahtani (southwest Arabia) which show that indeed the lateral fricative is present in these dialects but the data suggest that although both sounds can be found (the lateral fricative & the interdental) the speakers do not in fact reproduce the same 'split' as in Standard Arabic, but rather use both sounds in lexemes which have etymological Daad as well as in lexemes which have etymological Dhaa, i.e. these sounds are treated as allophones (rather than separate phonemes).

The analysis is based on research recently completed in southwest Arabia in two Qahtani Tihami locations using sociolinguistic interviews with 24 native speakers of the dialect, distributed over two age and gender groups. The dialect of this region has not been studied in detail, and it was only mentioned in the descriptive work of Prochazka (1988). This area is of considerable interest particularly because the dialects spoken in the region preserves archaic Semitic/Arabic features, which in addition to a lateral pronunciation of Daad also include m-definite article and dropping of final consonants. Furthermore, the dialects of the region are interesting from a typological viewpoint as they are spoken at the periphery of Arabia.

**References**

- Al-Azraqi, Munira. 2010. The Ancient Dād in Southwest Saudi Arabia. *Arabic*, 57 (1). 57-67(11)
- Al-Wer, Enam. 2004. Variability reproduced: a variationist view of the Dhaa/Daad opposition in modern Arabic dialect. In M. Haak, R. De Jong & K. Versteegh (eds) *Approches to Arabic dialects*. Leiden: Brill.
- Corriente, Federico. 1978. D-L doublets in classical Arabic as evidence of the process of de-lateralisation of Daad and development of its standard reflex. *Journal of Semitic Studies* 23:1. 50-55
- Steiner, Richar. 1976. *The case for fricative laterals in Proto-Semitic*. New Have, Conn.: Americal Oriental Society.
- Prochazka, T (1988) *Saudi Arabian Dialects*. London: Kegan Paul International

**To Metathesize or Not to Metathesize: Sonority Constraints in Tunisian Arabic Nouns**  
*Suyeon Yun, Massachusetts Institute of Technology*

This paper aims to provide a precise description of the distribution of CVCC and CCVC nouns in Tunisian Arabic spoken in Tunis (TA) and a formal account of it within Optimality Theory (Prince and Smolensky 1993/2004). In TA, unstressed short vowels in open syllables are syncopated, and thus CVCVC words in Classical Arabic or other Tunisian dialects like Sfaxian appear as CCVC (e.g., /qalam/ → [qlam] ‘pen’) in isolation or before a consonant-initial suffix (while they become CVCC before a vowel-initial suffix). What interests us is that a considerable number of CVCC nouns undergo metathesis of the vowel and the following consonant, also resulting in the CCVC structure, as shown in (1). It has been suggested for Moroccan Arabic that the position of a vowel in C<sub>1</sub>C<sub>2</sub>C<sub>3</sub> roots is determined by relative sonority between C<sub>2</sub> and C<sub>3</sub>; (i) CvCC, if |C<sub>2</sub>| > |C<sub>3</sub>|, (ii) CCvC, if |C<sub>2</sub>| ≤ |C<sub>3</sub>| (|C| = relative sonority of C; Boudlal 2006). However, the generalization does not hold for the TA data as seen in the counterexamples (2). Although there have been a couple of previous studies (e.g. Wise 1983) that report metathetic change of CVCC into CCVC of a few verbs, no comprehensive study has been done about TA nouns.

Based on 268 CVCC or CCVC noun forms from Ben Abdelkader (1977) and from my own fieldwork, I argue that two basic principles play a key role in determining the CVCC or CCVC surface form. First, rising sonority in coda clusters is avoided, as in Moroccan, and is likely to be split by the metathesis of the preceding vowel explaining (1). This is consistent with the Sonority Sequencing Principle (Clements 1990 and many others) requiring sonority fall from the nucleus throughout the coda. Next, examples with sonority-rising coda in (2) are tolerated since they have corresponding outputs, usually /-a/-suffixed forms (feminine or singulative) as in (3), which are frequently used and with which the identity might be required. This can be also supported by the fact that, for example, even when there exists an singulative /-a/ form, [lahma], for collective [lham] ‘meat’, Tunisians do not use CVCC form before a vowel-initial suffix, e.g., [lahm-i] ‘meat’ + ‘my’ for ‘my meat’ but use [lham mta:ʕi] ‘meat’ + ‘mine’, maintaining the CCVC form, which indicates that other corresponding forms are much less available for [lham]. Still, there are examples not explained by these two principles, which may need some extra-phonological explanation.

The TA data is accounted for by interactions of LINEARITY constraints (McCarthy and Prince 1995) banning metathesis and a markedness constraint on coda clusters. LINEARITY-IO that prohibits metathesis of input segments in the output is ranked below CODA-SONORITY that prohibits rising sonority coda clusters, and thus metathesis occurs to avoid a sonority-rising coda cluster (4). On the other hand, LINEARITY-OO that prohibits metathesis of segments in another output form outranks CODA-SONORITY, which prefers to keep the CVCC form for the correspondence to a CVCC-a form (5). In sum, constraint ranking LINEARITY-OO >> CODA-SONORITY >> LINEARITY-IO can predict a significant portion of the distribution of TA monosyllabic nouns.

- (1) /lahm/ [lham] ‘meat’ (2) /fikr/ [fikr] ‘thinking’ (3) [fikr-a] ‘idea’  
 /Sahn/ [Shan] ‘plate’ /hukm/ [hukm] ‘judgment’ [hikm-a] ‘wisdom’  
 /ʕumr/ [ʕmur] ‘age’ /xubz/ [xubz] ‘bread’ [xubz-a] ‘a piece of bread’

(all examples in this abstract are from my field notes)

(4) /lahm/	LINEAR-OO	CODA-SONORITY	LINEAR-IO	(5) /fikr/	LINEAR-OO	CODA-SONORITY	LINEAR-IO
[lahm]		*!		→ [fikr]		*	
→ [lham]			*	[fikr]	*!		*

**Selected References**

Ben Abdelkader, R. 1977. *Peace Corps English-Tunisian Arabic Dictionary*. Peace Corps, Washington D.C. // Boudlal, A. 2006. Sonority-driven schwa epenthesis in Moroccan Arabic. *Languages and Linguistics* 18. 59-81. // Wise, H.1983. Some functionally motivated rules in Tunisian phonology. *Journal of Linguistics* 19(1). 165-181.

**Session: Thursday, 11:45 AM ~ Keynote Address**

**Explorations at the syntax-phonology interface in Arabic**  
*Sam Hellmuth, University of York, UK*

Despite an abundance of research on Arabic syntax and phonology as separate domains, there is as yet relatively little research at the syntax-phonology interface in Arabic. This talk begins by providing an overview of what we know so far, in an effort to identify reasons for the lack of work at the interface to date. It is possible that research in this domain has been held back by the lack of comparable descriptions of the prosody of spoken dialects.

Redressing this lack is a key goal of the Intonational Variation in Arabic project ([www.york.ac.uk/res/ivar](http://www.york.ac.uk/res/ivar)). The talk concludes with a preview of two case studies of the potential benefits that directly parallel prosodic descriptions of Arabic dialects will bring to our understanding of: i) the typology of basic phrasing patterns, and ii) the interaction of prosody with negation, wh-interrogatives and focus, in spoken Arabic dialects.

**Session: Thursday, 1:45 PM**

**On the Featural Properties of Complementizers in Arabic**

*Hamid Ouali, University of Wisconsin-Milwaukee*

On the featural properties of Complementizers in Arabic Proposal: In this paper, we analyze the featural properties of the different Complementizers in Arabic namely  $\text{?an}$  (1) on one hand, and  $\text{?anna}$  (2) and  $\text{?inna}$  (3) on the other hand in light of Chomsky's (2007, 2008, and 2012) C-T feature-inheritance approach. We will propose that  $\text{?an}$  is lexically specified for Phi-features whereas  $\text{?inna}$  and  $\text{?anna}$  are specified for Phi and Tense features (see Fassi-Fehri 2012 for C and Tense). The featural specifications of  $\text{?an}$  and  $\text{?anna}/\text{?inna}$  explain the agreement and tense properties of the clauses that they each head.  $\text{?an}$  clauses (such as ECM constructions) are tenseless but exhibit subject-verb agreement (unlike English ECM constructions) and  $\text{?anna}$  and  $\text{?inna}$  clauses show both tense and subject-verb agreement.

Theoretical Assumptions: Chomsky (2007, 2008, and 2012) has proposed a feature inheritance analysis to explain the difference between the distribution of bare infinitival TPs (4) and finite TPs (5), which was a mere stipulation in GB theory. Under Chomsky's analysis, T in the lexicon uniformly lacks phi-features (and tense for that matter), while C bears them. Empirical evidence comes from languages such as Dutch and West Flemish which exhibit C-agreement.

If T is selected by C, then T can inherit C's Phi-features by the mechanism of feature inheritance. Under this analysis, the morphological features of certain functional heads (here, C and T) change in the course of the derivation, but not in the lexicon. This suggests that the cross linguistic parametric variation might be a result of two factors. One is the featural properties of function heads in the lexicon (see e.g. Benmamoun 2000) and two is whether the conditions for feature-inheritance are available for Agree to take place.

Analysis:  $\text{?an}$ , typically selected by want-type ECM verbs (1), is specified for Phi features. Upon merging with TP, C ( $\text{?an}$ ) gets its Phi-features inherited by T. T then establishes an Agree relation with the subject (6) and (7). The tenseless T also lacks the EPP feature hence the subject never occupies Spec, TP (8).  $\text{?inna}/\text{?anna}$  are specified for Phi and Tense features, which T inherits. This tensed T also carries an EPP feature (presumably inherited as well). All this explains why  $\text{?inna}/\text{?anna}$  clauses are tensed, show agreement, and their Spec,TP's must be occupied by a DP or pronoun (3) vs. (9).

This DP, however, does not internally merge/move to this position. We will argue that in both  $\text{?an}$  and  $\text{?inna}/\text{?anna}$  clauses the thematic subject never moves to Spec-TP. Spec-vP is uniformly the thematic and the Nominative Case position of the subject (Chomsky 2000 and subsequent work). If T carries an EPP feature, a DP argument must be merged in its Spec to check this feature. In English this DP is the subject and its placement is derived by internal merge/movement, in Arabic this DP

is externally merged (Soltan 2007) and could be a Topic (2)-(3) or a Left-Dislocated Object (10) (Fassi-Fehri 2012).

Implications: This analysis maintains the approach, first advanced within Principles and Parameters (Chomsky 1993 following Borer 1984), where syntactic parameters are restricted to variation in the morphological features of functional syntactic heads. In other words, according to this restrictive theory of parametric variation, functional categories in different I-languages may differ slightly in their morphological features, as represented in their lexical entries and these small differences give rise to radical differences in the 'appearance' of 'languages' (see e.g. Benmamoun 2000 and Aoun et al 2010 for Arabic).

**On the Syntax of Spatial 'Prepositions' in Lebanese Arabic**  
*Lina Choueiri, American University in Beirut*

In this paper, I examine the syntax of a class of locative 'prepositions' in Lebanese Arabic, which might be descriptively termed as 'lexical' or 'semi-prepositions' (Ryding 2005). Table (1) lists those expressions and the example in (2) illustrates their use.

(1) Locative semi-prepositions in Lebanese Arabic

*foʔ* 'above', *taħt* 'below', *ʔiddem* 'in front of',  
*wara*, *ħalf* 'behind', *barra* 'outside', *zuwwa*  
 'inside', *ħadd* 'near', *ʕind* 'at', *maʕ* 'with', *beyn*  
 'between'

(2) *wiʔif* *foʔ/taħt/ʔiddem/wara/ħalf/ħadd/barra(t)/zuwwa(t)* I-bineeye  
 stood.3sm above/below/in front of/behind/behind/near/outside/inside the-building  
 'He stood above/below/in front of/behind/behind/near/inside/outside the building.'

The syntax of Arabic prepositional phrases, in general, and locative prepositional phrases, in particular, has not been the subject of much discussion in the generative literature; therefore, this paper breaks new ground by documenting important aspects of the analysis of spatial 'prepositions' within that perspective. Specifically, adapting a structure initially proposed by Terzi (2006) for Greek, I argue that the 'semi-prepositions' in table (1) are projected in a complex constituent structure, schematically represented in (3).

(3) [<sub>PP,LOC</sub> P<sub>LOC</sub> [<sub>DP</sub> D [<sub>NP</sub> PLACE [<sub>DP/PP</sub> ]]]]

In (3), a functional head, P<sub>LOC</sub>, takes an obligatory nominal complement, whose head noun 'PLACE' can be silent/null. PLACE, in turn, can take either a DP or a PP complement. I further show that the 'semi-prepositions' in table (1) can be divided into two sub-classes: those that occur as prepositional heads in the PP complement of PLACE (P-locatives), and those that spell-out the 'PLACE' noun (N-locatives) (4).

(4) [<sub>PP,LOC</sub> P<sub>LOC</sub> [<sub>DP</sub> D [<sub>NP</sub> PLACE [<sub>PP</sub> P [<sub>DP</sub> ] ] ]]]]  
*foʔ, taħt, ʔiddem, wara, ħalf, ʕind, maʕ, beyn*  
*barra, zuwwa, ħadd*

The structures in (3) and (4) provide a straightforward account for a set of observations about locative 'semi-prepositions' in Lebanese Arabic. First, while N-locatives can occur 'intransitively' (5a), P-locatives cannot (5b).

(5) a. *ʔaʕad* *foʔ/taħt/ʔiddem/wara/ħalf/barra/zuwwa*  
 sat.3sm above/below/in front of/behind/behind/outside/inside  
 'He sat above/below/in front of/behind/outside/inside'  
 b.\* *ʔaʕad* *ʕind/maʕ/beyn*  
 sat.3sm at/with/between

As seen in (4), being prepositional heads, P-locatives will require a complement, since Lebanese Arabic does not allow intransitive prepositions. On the other hand, N-locatives are generated as nominal heads and, as such, they do not require complements. The nominal nature of N-locatives is also seen in the fact that they give rise to adjectival modifiers (6).

(6) I-beeb I-foʔaane/I-taħteene/I-ʔidmeene/I-warraane/I-ħalfeene/I-barraane/I-zuwweene  
 the-door the-above/the-below/the-front/the-back/the-back/the-outside/the-inside  
 'The (front/back/outside/inside) door (above/below)'

Another interesting observation that differentiates P-locatives from N-locatives is the fact that, while N-locatives can occur as the first members of a complex preposition with *min* 'from', P-locatives are not allowed in such constructions (7).

(7) a. *wiʔif* *foʔ/taħt/ʔiddem/wara/ħalf/ħadd/barrat/zuwwet* min-o  
 stood.3sm above/below/in front of/behind/behind/near/outside/zuweet from-it  
 'He stood above/below/in front of/behind/behind/near/outside/inside it.'  
 b.\* *wiʔif* *ʕind/maʕ/beyn* min-un  
 stood at/with/between from-them  
 'He stood at/with/between them.'

I take *min* 'from' to be a prepositional head. A schematic representation of the locative prepositional phrase in (7a) is given in (8).

(8) [<sub>PP,LOC</sub> P<sub>LOC</sub> [<sub>DP</sub> D [<sub>NP</sub> PLACE [<sub>PP</sub> *min* [<sub>DP -o</sub> ] ] ]]]]  
*foʔ, taħt, ʔiddem, wara, ħalf,*  
*barra, zuwwa, ħadd*

Thus, *min* heads the PP complement of N-locatives. The unacceptability of (7b) is then explained by the fact that *min* occupies the same position as P-locatives, and is therefore in complementary distribution with those prepositions.

The N-locative *ħadd* 'near' appears to behave differently from other N-locatives in not allowing a silent complement: (9) contrasts with (5a). However, like other N-locatives, *ħadd* 'near' can form a complex preposition with *min* (7a).

(9) \* *ʔaʕad* *ħadd*  
 sat.3sm next

I account for the singular behavior of *ħadd* 'near' by following Choueiri and Ouhalla (in preparation) in assuming that *ħadd* 'near' is obligatorily a relational noun, which requires an argument. In that sense, it behaves like inalienably possessed nouns. In my talk I will further develop the analysis of N-locatives and P-locatives and examine its cross-linguistic consequences.

Session: Thursday, 2:15 PM

On the Syntax of *ʔillaa* in Egyptian Arabic

Usama Soltan, Middlebury College

A typical exceptive construction in Egyptian Arabic (EA) is one headed by *ʔillaa* (=‘except’), followed by an XP, where XP may be a DP, PP, adverbial, or CP, as in (1-4), respectively.

- (1) ʔanaa maa-ʔaabil-t-i-š *ʔillaa* Ahmad  
 I neg-met-1sg-ev-neg except Ahmad  
 ‘I did not meet anyone except Ahmad.’

- (2) ʔanaa miš ha-xrug *ʔillaa* maʔa Ahmad (3) *ʔillaa* bukra  
 I neg fut-go.out.1sg except with Ahmad .... except tomorrow  
 ‘I will not go out except with Ahmad/except tomorrow.’

- (4) ʔanaa miš ha-xrug *ʔillaa* ʔizaa xaragat maʔa-naa  
 I neg fut-go.out.1sg except if went.out.2sgm with-us  
 ‘I will not go out except if you go out with us.’

In this paper, I discuss the syntax of exceptive constructions in EA focusing on the following three questions: (i) What environments does an *ʔillaa*-phrase occur in? (ii) What is the syntactic category of *ʔillaa*? (iii) What is the syntactic structure of exceptive constructions?

With regard to question (i), I argue that *ʔillaa* is a Negative Polarity Item, whose distribution is similar, but not identical, to that of N-words in the language. In particular, in addition to its occurrence in antimorphic contexts (i.e., the scope of negation and *without*), *ʔillaa* also occurs with quantified expressions, as in (5).

- (5) ʔanaa ʔaabil-t kull ʔil-Talaba *ʔillaa* Ahmad  
 I met-1sg all the-students except Ahmad  
 ‘I met all the students except Ahmad.’

Like the N-word *walaa*, *ʔillaa* sometimes contributes negative force, and sometimes does not. A syntactic agreement analysis of negative concord in natural language (e.g., as in Zeijlstra 2008) is argued to account for the distribution *ʔillaa*.

As for question (ii), I argue that *ʔillaa* cannot be a preposition, since it is unable assign dative case to pronominals (\**ʔillaa-h*=‘except him;’ nominative form of the pronoun has to be used *ʔillaa huwwa*=‘except he’). Also, since the excepted XP can be a PP, treating *ʔillaa* as a P is rather problematic because a P selecting PPs is not a productive pattern in the language. Similarly, there is strong evidence that *ʔillaa* cannot be a focal adverb (such as *hatta*=‘even,’ or *bass*=‘only’). Unlike focal adverbs, *ʔillaa* cannot occur following the exceptive XP (\**Ahmad ʔillaa*; cf. *Ahmad hatta* or *hatta Ahmad*), and while multiple focal adverbs are allowed, multiplicity of *ʔillaa*-phrases is not.

Instead, to answer question (iii), I argue, following work by Hoeksema (1995) and Pérez-Jiménez and Moreno-Quibén (2012), on English and Spanish exceptives, respectively, that *ʔillaa* is a conjunction particle heading a Boolean Phrase (BP) whose second conjunct is a CP in which the excepted XP moves to the left periphery followed by TP-ellipsis, thereby giving rise to the surface exceptive structure, as shown in the syntactic representation in (6) for the sentence in (5).

- (6) [<sub>CP1</sub> ʔanaa ʔaabil-t kull ʔil-Talaba [<sub>BP</sub> *ʔillaa* [<sub>CP2</sub> Ahmad<sub>i</sub> [<sub>TP</sub> ʔanaa ʔaabil-t<sub>i</sub>]]]]

Evidence for an underlying coordination structure comes from effects of the Coordinate Structure Constraint with regard to scope freezing. In particular, structures that are scopally ambiguous lose this ambiguity once an exceptive phrase occurs. Further evidence comes from exceptive constructions with multiple remnant XPs. Differences between Spanish and EA in the syntax of exceptive constructions are, then, argued to follow from certain restrictions on movement to the clausal left-periphery in EA.

### **References**

- Hoeksema, J. 1995. The semantics of exception phrases. In *Quantifiers, Logic and Language*, ed. J. van der Does and J. van Eijk, 145–177. Stanford: CSLI Publications.
- Pérez-Jiménez, I. and N. Moreno-Quibén. 2012. On the syntax of exceptions. Evidence from Spanish. *Lingua* 122: 582-607.

## Session: Thursday, 3:30 PM

### The Semantic Complexity of Kuwaiti Arabic: The case of *dašš*

*Yousuf AlBader, University of Sheffield*

The study reports on an investigation of the lexical behavior of the Kuwaiti Arabic (KA) verb *dašš* 'to enter'. This verb is among the most frequently used verbs in everyday speech (cf. Holes 2001, Qafisheh 1999). It is prototypically a verb of change of location which entails that the moving entity changes location during the process. But this study shows that *dašš* is one of the polysemous verbs that can be metaphorically extended with more than twenty senses. For instance:

- |   |                                       |
|---|---------------------------------------|
| 1. <i>il-alwan dašša fi ba'zha</i>        | (The colors blend one into the other) |
| 2. <i>Noora dašš-at šahar-ha it-tasi'</i> | (Noora went into labour)              |
| 3. <i>Noora dašš-at galbii</i>            | (I am falling in love with Noora)     |
| 4. <i>aħmad dašš b-ğaybooba</i>           | (Ahmad fell into a coma)              |

My analysis makes the following claims: (i) the lexeme *dašš* is a polysemous word, and (ii) it can be lexically distinguished from its near-synonym *daxal*. The KA data acquired for this study were collected via questionnaires, personal interviews, and Arabic/Gulf Arabic dictionary entries. The variety of KA described in this study is the urban Kuwaiti speech.

Four dictionaries were consulted to trace the historical changes in meaning of *dašš*. The Gulf Arabic dictionaries were: Holes (2001) who lists eight senses for *dašš* including the causative form *daššaš* and the passive *indašš* and some metaphorical senses, whereas Qafisheh (1999) lists three only. But there is no mention of the verb in the modern Arabic dictionaries (cf. Buckwalter and Parkinson 2011, Wehr 1993).

The questionnaire results showed that for native speakers of KA, *dašš* is slightly different from *daxal* and expresses certain attributes. For example, *dašš* carries the meaning 'to enter with sudden impetuosity' such as: *il-ħaraamii dašš bayt-na* 'the thief quickly/suddenly barged into/broke into our house', as compared to *il-ħaraamii daxal bayt-na* 'the thief entered our house'. *Dašš* could mean 'the thief broke into the house through the window or the chimney', while *daxal* indicates 'the thief entered through the door'. On the other hand, *dašš* could imply that the motion event happens unintentionally.

This can be demonstrated in the sentence: *is-sayyara dašš-at feeni* 'I had a car accident (lit The car entered me)' which KA native speakers rated higher than *is-sayyara i-daxal-at feeni* because the collision occurred by accident. Basically, each verb selects its preferred complement and meaning varies with context. My study concludes that *dašš* is in fact a productive and polysemous word, and the analysis of these diverse data show: (i) the verb *dašš* is a member of the domain of motion verbs, (ii) several lexical and grammatical meanings can be expressed by one and the same lexeme, (iii) we have seen some possible denotational ranges of this single word, (iv) it expresses several meanings in terms of a semantic motivation, in this

case 'polysemy', (v) and finally this linguistic entity becomes conventionalized on the level of the lexicon, which can be distinguished from its near-synonym *daxal*.

#### References

Buckwalter, Tim and Parkinson, Dilworth. 2011. *A Frequency Dictionary of Arabic: Core Vocabulary for Learners* (UK: Routledge)

Holes, Clive D. 2001. *Dialect, Culture and Society in Eastern Arabia, Volume I: Glossary*. (Leiden, Boston, Cologne: Brill)

Qafisheh, Hamdi A. 1999. *NTC's Gulf Arabic-English Dictionary*. (USA: McGraw-Hill Contemporary)

Wehr, H. 1994. *Dictionary of Modern Written Arabic: Arabic-English*. (USA: Spoken Language Services Inc.)

**Session: Thursday, 3:30 PM**

**Identifying Semantic Relations in Arabic**

*Sameh Alansary, Bibliotheca Alexandria*

Semantic roles were introduced in 1960s and 1970s as a way of classifying the arguments of natural language predicates into a closed set of participant types which were thought to have a special status. The general idea of semantic roles has played an important part in linguistic theory in the past twenty years or so as they play a central role in language comprehension. There are many acknowledged attempts for the semantic analysis of natural languages such as the FrameNet and the Propbank. Few approaches are devoted to Arabic.

One of the promising systems that contain a list of semantic roles for representing the relations between words in sentences of different natural languages is the Universal Networking Language (UNL). Relations constitute one of the major components of UNL. The UNL has about 43 relations that have been normally used to represent semantic cases or thematic roles between concepts. These relations are represented as two-character or three-character lower-case strings such as agt, obj, plc, scn, tim, gol, ....etc. In the UNL framework, relations describe the semantic link between two concepts. These links are binary and directed (from a source to a target) and are claimed to be universal. Because of their similarity in name and function to syntactic relations, it may seem that these labels are different names for special grammatical functions. This is emphatically not the case. The intention is that these labels denote specific ideas rather than grammatical structures: the idea of “something that initiates an event,” or “agent” for example, is quite different from the “grammatical subject of a sentence”, even though many times the subject of a sentence will indicate the agent of the event. The agent of an event may also appear as an adjective or noun modifier, preceded by preposition “by” or embedded in nouns with the suffix “er”. The whole point of conceptual relations is to have a name for these very different grammatical structures which are conceptually quite the same. Thus, the conceptual relations used in UNL are much more abstract than the known grammatical relations.

This paper aims at classifying semantic relations in Arabic using the set of relations identified in the UNL framework. And how can use this classification to build a semantically annotated Arabic sentences and many other applications and resources for natural language processing. If we use the Arabic sentence in (1) as an example, النيل قلادة يعقوب مجدي الدكتور مصر منحت, - ‘Egypt has granted Dr. Magdy Yacoub the Order of the Nile.’

The UNL can identify the semantic relations between the words in this sentence as the following: the main predicate of the sentence is the verb “منح”, ‘granted’, there is an “agt” relation between “منح” ‘granted’ and the subject “مصر” ‘Egypt’, an “obj” relation between the same verb “منح” ‘granted’ and “النيل قلادة” ‘Order of the Nile’

and “gol” relation between the verb “ منح ” ‘granted’ and “ يعقوب مجدي الدكتور ” ‘Dr. Magdy Yacoub’.

Also, in the noun phrase “ يعقوب مجدي الدكتور ” ‘Dr. Magdy Yacoub’, there is a “nam” relation between “ الدكتور ” ‘Dr.’ and “ يعقوب مجدي ” ‘Magdy Yacoub’. This classification is represented in the UNL as in figures (1) and (2).

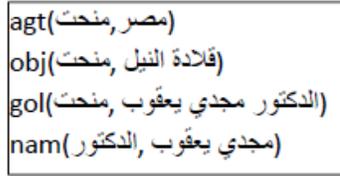


Figure (1)

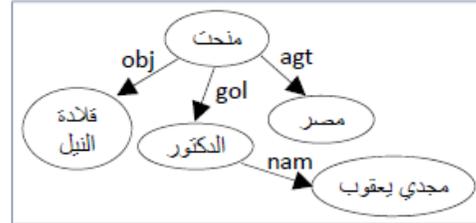


Figure (2)

The set of UNL relations includes many other semantic relations that can represent the same syntactic structure but different meanings according to the nature of the verb. Take the examples in (2), (3) and (4).

- (2) “قال محمد الحقيقة”, ‘Mohamed told the truth’
- (3) “عرف محمد الحقيقة”, ‘Mohamed knew the truth’
- (4) “ظهرت الحقيقة”, ‘the truth appeared’

The semantic relation between the verb and its specifier differs in the three sentences according to the type of the verb. In sentence (2) the verb “ قال ” ‘told’ is a do verb and its subject is the initiator of the action so the relation between “ قال ” ‘told’ and “ محمد ” ‘Mohamed’ is an “agt” while in sentence (3) the verb “ عرف ” ‘knew’ is a be verb so the relation between “ عرف ” ‘knew’ and “ محمد ” ‘Mohamed’ is an “aoj” which indicates a thing in a state or has an attribute, and finally in sentence (4) the verb “ ظهرت ” ‘appear’ is an occur verb (the subject of the verb is affected by the verb itself) so the relation between “ ظهرت ” ‘appeared’ and “ الحقيقة ” ‘truth’ is an “obj”. many other issues can be handled by the semantic relations in the UNL framework such as the representation of the prepositions meanings and so forth.

Depending on this identification of the different semantic relations in Arabic many applications and tools can be built such as semantic web applications, information retrieval systems, information extraction systems, machine translation systems and summarization systems. The semantic graph illustrated in figure (2) can be used to extract information or to delete the unessential nodes from the graph and produce a summarized graph that can be converted to a summarized natural language sentence as well as it can be used to translate it to any target language.

This paper will discuss in details the strength of the set of semantic relations of the UNL system in the semantic annotation of Arabic and the ability to deal with the various Arabic structures and issues related to the complexity of the Arabic structures.

**Session: Thursday, 4:45 PM ~ Keynote Address**

**Arabic Verbal And Nominal Plurals And The Syntax Morphology Interface**  
*Elabbas Benmamoun, University of Illinois, Urbana-Champaign*

One of the key assumptions of the Minimalist Program is that (narrow) syntactic derivations are driven by the interaction between formal features. However, the crucial aspect of formal features that drive their syntactic activity is the lack of interpretability at the LF interface. Thus, non-interpretable agreement features on temporal and verbal heads must be paired with interpretable matching features on nominal elements such as subjects and objects.

In this paper, I shall attempt to show that this dichotomy is correct and has far reaching consequences for the syntax morphology interface. In particular, it will be shown that number marking which can realize both verbal plurality (at the level of event structure) and number agreement with subjects and objects differ in how they are derived (and which prosodic units they can access) depending on whether they are interpretable or non-interpretable. The critical evidence comes from Arabic where plural marking realizing verbal plurality seems to access the sub-word level (possibly the root or prosodic units such as syllables and feet) but plural marking realizing subject agreement accesses the word level. This raises important questions about the interface between morphology, the lexicon, and syntax.

**Session: Friday, 8:30 AM**

**On the Clausal Anchoring of Parenthetical Verb Phrases in Moroccan Arabic**

*Noureddine Elouazizi, Simon Fraser University*

1. Background: Problem

This paper examines the syntax of parenthetical verb phrases (henceforth PVPs) such as: *tanden* “I think”, *taneCtaqed* “I believe” and *taneftard* “I suppose” in Moroccan Arabic (henceforth MA) with a special focus on their ordering properties and attachment to the clausal hosts. Moroccan Arabic is a pro-drop VSO language (Abdel-Massih 1973, Wager 1983, Benmamoun 1992, 2000, a.o.). I examine the properties of the PVPs in Moroccan Arabic in light of two questions: (i) are PVPs in Moroccan Arabic integrated or detached from their clausal hosts? and (ii) does the syntax of the PVPs affect or is affected by that of the host clause?

2. Previous accounts

The syntax of parentheticals has been examined in previous generative grammar studies. On the one hand, the syntax of parentheticals seems to exhibit syntactic dependencies with the clausal environments they occur in (See Ross (1973), Emonds (1976, 1979), McCawley (1989)). On the other hand, the syntax of parenthetical seems to be invisible to many syntactic constraints (See Espinal (1991), McCawley (1989), and de Vries (2007)) and exhibits a detachment “orphanage” from the host clause (Haegeman 1991). There is no unified view of the source structure of the PVP and the way its surface positions are derived. The syntactic properties of the PVPs in MA exhibit a set of mixed properties. In (1), PVPs break the adjacency between the subject and the verb and between the subject and the object positions in the VSO word order. In (2), the same flexible patterns are observed in structures of arguments promotion and demotion. This suggests that PVPs in Moroccan Arabic are detached from the internal clausal structures of their hosting clauses. However, the occurrences of the parenthetical verbs in the contexts of negation and polar questions indicate that they can actually be an integral part of the internal structure of the host clause (see examples 3 and 4).

3. Analysis

Adopting the Minimalist Program framework (Chomsky 2001, 2008), I propose that PVPs in Moroccan Arabic are reduced CP structures (with missing complement) and that they can be affected by the syntax of the host clause, namely by the configurational and interpretative properties of subjecthood in the host clause. I argue that for PVPs to be licensed they must be in the scope of a matrix clause referential subject, and establish a referential chain with such a subject position. Supporting evidence for this analysis includes:

First, the interaction of PVPs and negation in (3.a-b) suggests that it is possible to negate the PVP. Moroccan Arabic allows for negating the mitigation and the assertion together or negating the assertion in the presence of a mitigating PVP verb, all depending on where the PVPs are inserted in the clausal structure. Second, the examples in (5.a-c) show that PVPs cannot occur between an indefinite subject and an auxiliary (5.c). The licensing of the PVP requires a referential matrix subject. Third, in the context of control, it is not possible for the PVP to occur in the syntactic domain of the controlled argument position, (See example 7.a&b and 8.b). The PVP is “parasitic” on the selectional and thematic relations that hold in the control configurations. The Spec position of the PVP hosts a speaker oriented subject “I”, making PVP a speaker oriented clause. Since PVPs update the meaning of the host clause, the two referential chains of the subject in PVP and the subject in control constructions intersect and the subject in PVP is parasitic on the position of PRO. This establishes two main conclusions. First, PVPs in MA can be affected by the syntax of the host clause (in control configurations, context of negation) and as such they are not “orphaned” structures. Second, for PVPs to be licensed they must not be in the scope of a non-referential indefinite subject.

If this line of reasoning is on the right track, then we should find instances of the occurrences of the parentheticals that exhibit similar constraints in genetically unrelated languages; namely that the parentheticals can be licensed in the scope of referential subjects but not in the scope of non-referring/weak referring subjects. This is exactly what we find in Italian. Cardinaletti (1994b:71) reports that weak (indefinite) *wh*-phrases such as Italian *che* cannot be followed by parentheticals, whereas strong *wh*-phrases can, as illustrated by the examples in (9.a&b). This much accounts for the structural and configurational properties of PVPs in MA. To derive these properties, we assume along the lines of Chomsky (2008: 139) that Lexical Items (LIs) are endowed with an Edge Feature that enables them to merge and to iterate without limit. When two lexical items, say LI1EF and LI2EF with two active Edge Features, are merged, the Edge Feature LI2 is eliminated by transferring LI2 to the interfaces in the spirit of multiple spell-out type of derivation (Epstein et al 1999). Being derived syntactic objects (with a discharged/eliminated Edge Feature), PVPs can be merged only to the edge of their host clause and they will never project (on their own). The resulting projection would always be that of their host. This accounts for why PVPs (as adjuncts) are never part of the internal Head-Complement relations/domain but they can and actually must be part of the ‘modificational’ relations that involve Specs (subject positions).

#### 4. Conclusion

The analysis of the configurational and the distributional properties of the PVPs in MA shows that PVPs: (i) do not feed or bleed the linear order in the host structure, (ii) they feed (in a supplementing way) the interpretation of the syntactic object they attach to (they mitigate the assertion of the matrix verb), and (iii) they bleed configurational and interpretative chains in some of the subject-changing transformations.

## 5. Data

- (1) kkteb *tanden* ttalib (*tanden*) had lktab (*tanden*) f London  
 write I think student I think this book I think in London  
 “A student wrote, I think, this book in London.” (Moroccan Arabic)
- (2) (*tanden*) had lktab (*tanden*) te-kkteb (*tanden*) f London (*tanden*)  
 I think this book I think PASS-write.PART I think in London I think  
 “This book was written in London.” (Moroccan Arabic)
- (3) a. ttyara *ma tanden sh* ghadi ttwsel f lweqt  
 plane NEG I think NEG FUT arrive in time  
 “The plane, I do not think, it will arrive on time.”  
 b. ttyara *tanden ma ghadi sh* ttwsel f lweqt (Moroccan Arabic)
- (4) Wash wll-a l-ddar?  
 wh return.PERF-3SM to-house  
 “Did he return home?” (Moroccan Arabic)
- a) nCam/wah (Yes) b) lla (No)  
 c) \* tanden (I think) d) (tanden) wll-a (tanden) l-ddar (I think he returned home)
- (5) a. Kayn ?idrab f rrbat  
 Be strike in Rabat  
 “There is a strike in Rabat.”  
 b. Kayn ?idrab *tanden* f rrbat  
 c. \*/?? Kayn *tanden* ?idrab f rrbat (Moroccan Arabic)
- (6) a. Hawel Omar<sub>i</sub> (bash) [<sub><\_></sub><sub>i</sub> y-ssreq meni lflus]  
 Try-PERF Omar that 3SM-steal-PERF from-me money  
 “Omar tried to steal money from me.”  
 b. Omar qeneC lmudir<sub>i</sub> (bash) [<sub><\_></sub><sub>i</sub> y-zid lu f lmanda]  
 Omar convince.PERF director that raise him in salary  
 “Omar convinced the director to raise his salary.” (Moroccan Arabic)
- (7) a. \* Hawel Omar<sub>i</sub> (bash) [<sub><\_></sub><sub>i</sub> *tanden* y-ssreq meni lflus]  
 Try-PERF Omar that I think 3SM-steal-PERF from-me money  
 “Omar tried to steal money from me.”  
 b. \* Omar qeneC lmudir<sub>i</sub> (bash) [<sub><\_></sub><sub>i</sub> *tanden* y-zid lu f lmanda]  
 Omar convince.PERF director that I think raise him in salary  
 “Omar convinced the director to raise his salary.” (Moroccan Arabic)
- (8) a. Lbarlamani ttleb mn ttujar<sub>i</sub> (bash) [<sub><\_></sub><sub>i</sub> imulu-h]  
 MP appeal.PERF from traders that fund-him  
 “MP appealed to the traders to fund him.”  
 b. \* Lbarlamani ttleb mn ttujar<sub>i</sub> (bash) [<sub><\_></sub><sub>i</sub> *tanden* imulu-h]  
 MP appeal.PERF from traders that I think fund-him  
 “MP appealed to the traders to fund him.” (Moroccan Arabic)
- (9) a. Che, secondo te, deve fare?  
 b. Che cosa, secondo te, deve fare?  
 “What, according to you, [he] must do?” (Italian: Source-Cardinaletti (1994b: 71))

**Attitude Datives in Lebanese Arabic: Pronouns that merge too high to be bound**  
*Youssef Haddad, University of Florida*

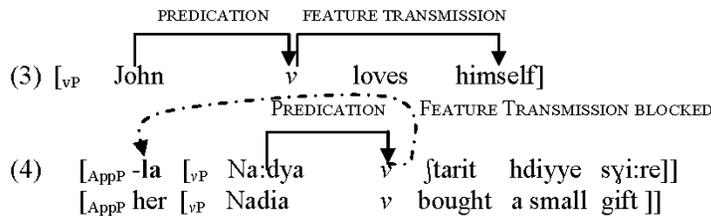
OVERVIEW: Lebanese Arabic licenses Attitude Dative (AD) constructions like (1) and (2). These contain non-truth conditional dative pronouns whose function is to express the speaker’s attitude – positive or negative – towards an event. An AD may refer to the subject, (1), or to the speaker or hearer, (2).

- (1) Na:dya {tarit-la sayyara zyi:re la-ʔibn-a  
 Nadia bought-her.DAT car small for-son-her  
 ‘Nadia bought a small car for her son. *Nothing significant.*’
- (2) Ziya:d byisʔrif-li/lak kil maʕa:ʃ-o ʕa-ʔakl l-hawa:  
 Ziad spends-me.DAT/you.DAT all salary-his on-eating the-air  
 ‘Ziad spends all his salary on nonsense stuff. *This is unacceptable.*’

The primary focus is on subject-oriented ADs; these seem to violate Condition B without leading to ungrammaticality. To account for this violation, first I show that ADs are high applicatives (Pylkkänen 2008). Then I explore two solutions: BINDING BY A FUNCTIONAL HEAD (Kratzer 2009) and MOVEMENT AND ANTI-LOCALITY (Grohmann 2003); I show that the latter is superior. Finally, I explain how ADs in general (i.e., ADs in structures like both (1) and (2)) determine their referent as subject or speaker/ hearer.

ANALYSIS: ADs do not belong to the thematic grid of the predicate; e.g., they are not beneficiaries or recipients, as (1) illustrates: the son is. This indicates that they merge as high applicative heads above vP.

Concerning the binding problem, (1) is considered a Condition B violation under the assumption that subjects are binders. However, Kratzer (2009) argues that little *v* and C are the true syntactic binders for pronouns. As (3) illustrates, a pronoun is bound by the closest *v* (or C) via two operations: PREDICATION (Specifier-Head Agreement) and FEATURE TRANSMISSION UNDER BINDING. As (4) shows, ADs merge above vP. Thus, they are not c-commanded by the syntactic binder *v*. This is why they are not reflexive.



Observe (5). The AD has an antecedent – a quantifier – in Spec,CP. This makes C a syntactic binder, yet no Condition B violation is induced. Data like (5) pose a problem for Kratzer’s approach to binding.

- (5) [CP kil wa:had C [IP nam-lo nisʔ se:ʕa baʕd l-ʕada:]]  
 [CP everyone C [IP slept-him.DAT half hour after lunch]  
 ‘Everyone slept for a half hour after lunch.’

An alternative is the movement approach. Grohmann (2003) argues that a mono-clausal sentence contains three Prolific Domains:  $\Theta$ -DOMAIN (vP) –  $\Phi$ -DOMAIN (IP) –  $\Omega$ -DOMAIN (CP). All licit movement takes place *across* Prolific Domains. Movement *within* a Prolific Domain is illicit unless the lower copy is pronounced. Grohmann further argues that reflexive pronouns are the result of movement within the  $\Theta$ -Domain, (6); they are copy spell-outs that are inserted in the derivation to save the structure.

- (6) a. [vP Sue [vP loves Sue]]      b. [vP Sue [vP loves herself]]

Unlike reflexive pronouns, ADs merge as high applicatives in the  $\Phi$ -Domain. They are not related to their antecedent via movement. Therefore, they are not subject to Condition B: No violation is induced.

Finally, I place subject-oriented ADs like (1) in the broader context of ADs; i.e., with speaker/hearer-oriented ADs like (2). I suggest that ADs determine their referent through Accessibility (Ariel 1988, 2001) by referring to a salient discourse topic. Subjects make good candidates due to their topic-like quality (Rizzi & Shlonsky 2007). Other salient discourse elements are speaker and hearer (Bhat 2004).



Session: Friday, 10:15 AM

**A Saliency-based Analysis of the Tunisian Arabic Demonstrative *hak*  
as Used in Oral Narratives**

*Amel Khalfaoui, Florida Atlantic University*

A speaker's choice in using different referring expressions in a language is not random; rather, according to the Givenness Hierarchy proposed by Gundel et al. (1993), different expressions signal differences in the assumed cognitive status (i.e. memory and attention status) of their referents. Within this framework, each cognitive status is a necessary and sufficient condition for the appropriate use of a different linguistic form or forms. Additionally, the statuses are implicationally related, so a given form can be used for referents that meet the minimal required status, and it can also be used to encode higher statuses. The interaction of the Givenness Hierarchy with Relevance Theory (Sperber and Wilson 1986/1995), however, explains why speakers typically do not use forms associated with a lower status for reference to a discourse entity with higher status; a more restrictive form provides more specific information about the referent's cognitive status, and would therefore require less processing effort for their hearers (see Gundel and Mulkern 1998). When a less restrictive form is used, the expectation is that some additional contextual effects are to be derived in return for the extra effort. This paper gives evidence from Tunisian Arabic (TA) oral narratives showing that the use of a demonstrative for reference to a discourse entity exploits this expectation for the purpose of signaling the relative saliency of the referent. In doing so, the paper also contributes support to research arguing that linguistic expressions specialize in genre-specific discourse functions (e.g. Biber 1984, OH 2003).

Of particular interest to this study is the demonstrative determiner *hak*, determined in a previous study (Khalifaoui 2007) to signal Familiar status (discourse entity retrievable from memory or mentioned earlier in the discourse). Distribution of this form shows that while *hak* is never used for reference to an entity whose cognitive status is higher than Familiar across a variety of written and spoken genres, it is frequently used in folk tales to refer to an entity that is Activated (i.e. in current working memory) or even In Focus (the current focus of attention). A corpus of 34 TA folk tales is examined to investigate the factors motivating the use of *hak* for reference to Activated and In Focus entities. Results indicate that *hak* is typically used for an entity whose cognitive status is higher than in order to promote its referent to the most salient position in the discourse relative to other entities. Specifically, narrators use *hak* in order to redirect the attention of the hearer to a new topic in the discourse, or to an entity that will influence the events of the story, although it is not established as a new topic in the discourse. By using *hak* instead of a more restrictive form, the speaker instructs the hearer to derive extra contextual effects about the referent's status in the discourse, in line with previous studies showing how linguistic expressions are used to signal the relative saliency of discourse referents (Mulkern 2003, Brustad 2000, Baker 1995, OH 2003). In the example in (1), the referent of *hak l- qsar* (that palace) could be assumed to be

activated by being mentioned in the immediately preceding sentence. By using the demonstrative *hak* instead of a more restrictive form such as a demonstrative pronoun, the storyteller is redirecting the hearer’s attention to the palace, establishing it as the new topic of the discourse, and inviting him/her to derive additional contextual effects, such as the uniqueness of the palace, and the assumption that it will be the location for a set of important events that are going to happen.

(1)... bu-ha      ħatt-ha              fi    **qsar**    taħt    l- qāʕa  
           father-her    put.past.3ms -her    in    palace    under    the-ground

**hāk**    l- **qsar**      ħatta    ħad    fi    l-dinja      ma      ja-ʕrf-u    wiin  
       that    the-palace    Neg    one    in    the world    Neg    3ms-know    where

ʕā                      wa    illa              mniin      ju-sil              l-u..  
 come.past.3ms    and    or              where.from    3ms-arrive    to-it..

“...her father put her in **an underground palace**, and **that palace**, nobody in the world knows where it is located or from where **it** could be reached..”

**Session: Friday, 10:45 AM**

**Converging Linguistic Evidence: The synonymy of Arabic COME verbs**

*Dana Abdulrahim and Antti Arppe, University of Alberta*

Previous quantitative, corpus-based research on near-synonymy concerning COME verbs in Modern Standard Arabic has shown that the interaction of a wide range of constructional elements can explain the behavioral differences between four alternative verbs: *atā*, *ā'a*, *qadima* and *a ara* (Abdulrahim, in prep.). The present study re-examines the case of near-synonymy among these MSA COME verbs by contrasting two sources of linguistic evidence: (1) corpus-based based data on natural usage and (2) experimental data on production/selection, using a forced-choice task. This allows us to assess the extent to which context-based regularities apparent in the written form overlap with the collective intuitions of native speakers of a modern Arabic dialect.

For the corpus-based component of this study, we sampled 500 corpus hits for each of the COME verbs from the MSA component of *arabiCorpus* ([arabicorpus.byu.edu](http://arabicorpus.byu.edu)). These overall 2,000 sentences were annotated for various morpho-syntactic and semantic features characterizing the constructions that host each of the four verbs. This data frame is subsequently subjected to polytomous logistic regression analysis using the one-vs-rest heuristic (Arppe 2012). This statistical method allows us to see which contextual features are distinct for each COME verb (e.g. PRESENT TENSE for *atā*), the sum of which can be interpreted to represent a prototype characterizing each COME verb. Another benefit of running polytomous logistic regression on this data frame is the resulting probability estimates for every individual annotated corpus hit as well as the three other alternatives in all contexts. This can help us extract the most exemplary corpus sentences/uses of each COME verb, as well as those contexts in which the COME verbs appear to be interchangeable, i.e. potentially truly synonymous.

This corpus-based analysis was paired with experimental data from a forced-choice task that was completed by 30 literate Bahraini native speakers of Arabic. In this experiment, the informants had to read a number of sentences in MSA and choose the missing verb from a given list of verbs. The 50 experimental stimuli were chosen to represent the full breadth of contextual richness apparent in the corpus data and the entire diversity of probability distributions given a context, ranging from (1) near-categorical preferences of only one COME verb, through (2) a decreasing degree of preference of one of the COME verbs and an increasing though smaller likelihood of one or more of the alternatives, to (3) approximately equal probability estimates for all four COME verbs. The proportions of selected verbs per each of the 50 distinct contexts incorporated in the stimuli were compared with the matching corpus-based probability estimates, exhibiting a high and significant correlation ( $r_{\text{pearson}} = 0.747$ ;  $p < 2.2e-16$ ). A mixed effects logistic regression model was then fit with the individual forced-choice verb selections as the dependant variable and the contexts incorporated in the stimuli as independent fixed effects variables, and

participant as the lone random effect. The estimated verb-specific odds for the contextual predictors in this mixed model based on forced-choice selections were found to largely agree in both direction and strength with those of the corpus based model.

In conclusion, our analysis shows that the collective intuitions of the native Bahraini speakers of Arabic concerning the context-based associations governing the selection of the four COME verbs corresponded substantially with those that we had observed modeling the written corpus sample of MSA. While the intuition-based data validates our selection of the explanatory variables used in the corpus-based model, the comparison of two sources of data in such a study aims to examine the extent to which literate Arabic speakers (from a certain dialectal background) have internalized the rules and conventions of lexical uses in MSA.

#### REFERENCES

- Abdulrahim, D. (in prep.). "Quantitative approaches to analyzing COME constructions in Modern Standard Arabic". To appear in Hardie, A. and McEnery, T. (eds.), *Arabic Corpus Linguistics*.
- Arppe, A. 2008. Univariate, bivariate and multivariate methods in corpus-based lexicography – a study of synonymy. Publications of the Department of General Linguistics, University of Helsinki, No. 44.
- Arppe, A. 2012. polytomous: Polytomous logistic regression for fixed and mixed effects. R package version 0.1.4.
- Arppe, A. and J. Järvikivi (2007). Every method counts – Combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory*, 3:2, 131-159.

**Session: Friday, 11:15 AM ~ Keynote Address**

**Modeling Sociopragmatic Language Use in Social Media in Arabic and English:  
A Comparative Computational Perspective**

*Mona Diab, George Washington University*

Social media language is a treasure trove for mining and understanding human interactions. In discussion fora, people naturally form groups and subgroups aligning along points of consensus and contention. These subgroup formations are quite nuanced as people could agree on some topic such as liking the movie the matrix, but some within that group might disagree on rating the acting skills of Keanu Reeves. Languages manifest these alignments exploiting interesting sociolinguistic devices in different ways.

In this talk, I will present our work on subgroup modeling and detection in both Arabic and English social media language. I will share with you our experiences with modeling both explicit and implicit attitude using high and low dimensional feature modeling. This work is the beginning of an interesting exploration into the realm of building computational models of some aspects of the sociopragmatics of human language with the hopes that this research could lead to a better understanding of human interaction.

**Session: Friday, 1:30 PM**

**Computational Measures of Linguistic Variation**

*Mahmoud Abunasser and Elabbas Benmamoun,  
University of Illinois at Urbana-Champaign*

Mahmoud Abunasser and Elabbas Benmamoun, University of Illinois at Urbana-Champaign Measures of linguistic variation, also called linguistic distances, is one of the prominent topics in the growing field of dialectometry, which is concerned with quantifying linguistic differences and similarities and, often, links it to geographical distances between the areas where the relevant languages are spoken (Nerbonne and Kretzschmar, 2003). In this presentation, I will report on an on-going project developing computational linguistic variation metrics quantifying the lexical, pronunciation, and morphosyntactic distance between three varieties of Arabic from different regions of the Arabic speaking world. This research benefits greatly from Arabic dialectological research which has provided rich and highly detailed descriptions of variation across the different regions where Arabic is spoken, including areas where Arabic is not the majority language. It provides a framework for combining insights from traditional typological and dialectology research and computational linguistics with specific focus on Arabic.

The varieties under consideration are Gulf Arabic (GA), Egyptian Arabic (EA), and Moroccan Arabic (MA) represented by elicitations from speakers from Al-Ain (UAE), Cairo (Egypt), and Rabat (Morocco). The distances are based on elicitations of data from native speakers of the varieties under consideration. The speakers (and their parents) were born and raised in the same city of consideration.

For the purpose of evaluating the lexical and pronunciation distance, the Swadesh list (207 items) was elicited and phonetically transcribed for all varieties. In the elicitation sessions, subjects are continually reminded to provide utterances in the variety they use in a totally informal settings, they are also given the appropriate context for the lexical items when an ambiguity is expected, and the researcher made some adaptations to the Swadesh list to make it consistent with Arabic varieties, such as cases of infinitive verbs in the Swadesh list are replaced by past tense verbs with third person masculine singular agreement, which is the simplest verb form that is found in dictionaries. The lexical distance is based on the number of cognate words in the list. The pronunciation distance is based on the Levenshtein distance – A variant of the minimum edit distance – between pair of words in the parallel lists. Levenshtein distance was introduced to computational linguistics by Kessler (1995), and subsequently used by other researchers.

The morphosyntactic distance depends on the variation in patterns of morphosyntactic agreement. To quantify the amount of variation we investigated different patterns of agreement depending on the number, gender and person of the subject. We elicited data for different classes of verbs (sound verb, geminate verb, three verbs where one of the root radicals is a glide, and a verb with two glides). In

addition to the verbal paradigms that expressed subject-verb agreement, we also extended the investigation to object and possessive clitics or bound pronouns. The morphosyntactic variation metric is based on pronunciation similarity between the conjugations of the verbs in each pair of varieties. The variation metric is penalized if the pronunciation of one form is close to another form, which results in asymmetry of the variation metric.

Based on Preliminary findings, the lexical and pronunciation metrics showed a closer distance between Egyptian and Gulf varieties while Moroccan turned out to be more distant from the former two. With regard to the morphosyntactic metric, the closest varieties were Egyptian and Gulf, and the most distant varieties were Moroccan and Gulf.

#### Selected References

- Kessler, B. (1995, March). Computational dialectology in Irish Gaelic. In Proceedings of the seventh conference on European chapter of the Association for Computational Linguistics (pp. 60-66). Morgan Kaufmann Publishers Inc..
- Nerbonne, J., & Kretzschmar, W. (2003). Introducing computational techniques in dialectometry. *Computers and the Humanities*, 37(3), 245-255.

Session: Friday, 2:00 PM

**Glottal-initial Verbs 'axaḍ/'akal in Contact Situations: Data from Amman Arabic**

*Enam Al-Wer, University of Essex, and Hanadi Ismail*

Glottal-initial verbs 'axaḍ/'akal in contact situations: data from Amman Arabic  
Conjugation of glottal-initial verbs such as 'axaḍ/'akal varies quite considerably across Arabic vernaculars. In the traditional central and northern Jordanian dialects (most of which are of the Hōrani Type), marking of 'person' agreement is morphologically neutralised in the imperfect of 1st pers sing and 3rd pers masc sing, rendering the ambiguous form bōkil 'I eat/he eats'. In Amman, the Jordanian dialects that have this system (e.g. the dialect of Salt) are in contact with Levantine varieties that mark agreement with 'person' in these forms; thus: bākul 'I eat' vs. byākul 'he eats'. The contact situation in Amman seems to have triggered change at two levels in the speech of Ammanis with a Jordanian dialectal heritage (as opposed to Palestinian heritage). Firstly, it triggers structural change by introducing a distinction between 1st pers and 3rd pers masc; in this development, the previously unmarked form bōkil is reallocated to the 3rd pers masc only, and a new form, bākul, is introduced for 1st pers. Secondly, it triggers a change in the phonetic form of the 3rd pers sing masc, bōkil > bōkul. The data show that while the structural (paradigmatic) development has been focused to a large extent the phonetic realisation of bōkil ~ bōkul ~ byākul is diffuse, showing a relatively high degree of variation.

The paper provides an analysis of this feature within the framework of variationist theory and the principles of new dialect formation. The data upon which the analysis is based come from 'the Amman project', a large scale research which investigates the formation of the city's dialect. The data were collected using the standard format of a 'sociolinguistic interview' mostly, and in some elicitation tasks. These are supplemented with data obtained from earlier research in various other localities in Jordan.

Session: Friday, 2:30 PM

**Cities of Migration:  
Urbanization and Linguistic Change in Jeddah**

*Aziza Al-Essa, King Abdulaziz University*

One of the sociolinguistic consequences of cityward migration and rapid urbanization is dialect contact and the development of linguistically diffuse speech communities where there are no clear linguistic norms (Kerswill 2004:30). Cities in the Arab world witnessed massive rural–urban migration in the second half of the twentieth century and the sociolinguistic impact of this in-migration is addressed in Miller (2007). In Saudi Arabia, development planning strategies of the Saudi government led to highly polarized growth of three major urban centers in the country, one of which is the city of Jeddah. To counteract this urban polarization, the government adopted strategies which encouraged the re-distribution of regional population to achieve equitable distribution of national wealth. The present paper aims at examining the linguistic change initiated by the internal migration of people from the central and southern regions of Saudi Arabia to the city of Jeddah.

The influx to Jeddah of people who speak dialects that are distinct from the local dialect has reshaped the sociolinguistic scene in the city. The native population in Jeddah speaks a mixed dialect which had been largely shaped by geopolitical and socio-religious factors, most importantly the external migration of different ethnic groups from outside the Arabian Peninsula in the past centuries. The recent internal migration has resulted in a linguistically diffuse community in Jeddah. My data show that the process of supraregionalisation (Hickey 2010), which entails the avoidance of salient linguistic features in the input varieties, is operating in the speech of 100 immigrants in Jeddah. While the regionally marked morphophonemic variants [-its] and [-iʃ] of the feminine suffix (-ik) are levelled out from the speech of second generation immigrants, the socially salient stop variants [t], [d] and [ʔ] which mark the local dialect of the city are acquired at a remarkably lower rate.

In this paper I will present the details of the analysis of the dialect leveling and diffusion of linguistic variants in the speech of second generation immigrants in Jeddah, and I will argue that the direction of linguistic change, i.e. selection and propagation of the competing variants, is determined by the degree of the immigrants' orientation to a supra-regional norm rather than the local norm.

## References

Hickey, Raymond. (2010). Supraregionalisation. In Laurel Brinton and Alexander Bergs (eds), *Historical Linguistics of English (HSK series)*. Berlin: Mouton de Gruyter.

Kerswill, Paul. (2004). Social dialectology/Sozialdialektologie. In Klaus Mattheier, Ulrich Ammon & Peter Trudgill (eds.), *Sociolinguistics/Soziolinguistik. An international handbook of the science of language and society*, 2nd edn., Vol 1. Berlin: De Gruyter.

Kerswill, Paul. (2006). Migration and language. In Klaus Mattheier, Ulrich Ammon & Peter Trudgill (eds). *Sociolinguistics/Soziolinguistik. An international handbook of the science of language and society*, 2nd edn. Vol 3. Berlin: De Gruyter.

Miller, C. (2007). Arabic Urban Vernaculars; development and change. In C. Miller, E. Al-Wer, D. Caubet and J. Watson (eds), *Arabic in the City: Issues in dialect contact and language variation*. Abdington:Routledge.

**Session: Friday, 3:00 PM**

3:00-3:30      **French Nouns in CS with Two Arabic dialects: A comparison of community norms**  
*Rebekah Post, University of Texas at Austin*

**Session: Friday, 3:45 PM ~ Keynote Address**

**The Maghreb-Mashreq language ideology and the politics of identity  
in a globalized Arab world**

*Atiqa Hachimi, University of Toronto, Scarborough*

To date, most scholarship on Arabic language ideologies has focused on the relation between Standard Arabic and the spoken vernaculars. This paper in contrast draws attention to the understudied relationship among the regional varieties of vernacular Arabic. Specifically, it seeks to make visible the workings of what it calls the 'Maghreb-Mashreq language ideology': the hierarchical relationship between Mashreqi (Middle Eastern) and Maghrebi (North African) vernacular Arabic varieties.

The paper explores in particular the de/authentication of Arabness through a detailed analysis of a transnational pan-Arab talent TV show. Drawing on clips of situated interactions from this series which have been uploaded to YouTube and commented upon by viewers, the paper argues that the new media is a critical site for reworking longstanding language ideologies and the politics of identity in the Arabic-speaking world.

**Session: Friday, 5:00 PM**

**The Antecedents of Arabic Dialects: A speech communities approach**

*Alexander Magidow, University of Texas at Austin*

Understanding the development of Arabic dialects has been hindered by their history of major population movements followed by lengthy periods of dialect contact and the diffusion of linguistic materials between dialects that are not necessarily historically related.

This paper argues that to understand the history of Arabic dialects, we need to move beyond the notion of ‘language’ to understand instead the history of ‘speech communities’ which can be multilingual or multidialectal. I define speech communities as sets social networks united by a sense of allegiance, such that the flow of linguistic features is bounded both by network connections (Milroy 2008) and by a speaker’s personal sense of allegiance with various social groupings, which they express by way of linguistic acts of identity (LePage and Tabouret-Keller 1985). These centripetal acts of identity direct linguistic change so that the boundaries of the expansion of a linguistic change are congruent with the boundaries of a speech community. I argue that this view allows us to more fully deploy social, historical and literary data than in the comparative method of historical reconstruction, while still using historical linguistic data to reconstruct earlier speech communities.

Since speech communities are defined both linguistically and socially, I begin with an overview of the sociohistorical situation in the pre-Islamic and early Islamic world from a linguistic perspective. I argue that the principle of First Effective Settlement (Labov 2001; Zelinsky 1992), the notion that the original settlement in an area determines the linguistic behavior of that area, allows us to look primarily at the earliest viable Arabic settlements, without a need to delve too deeply into later historical eras. I argue that based on this data, models which propose massive changes in Arabic as a result of second language learning (Versteegh 1984; al Sharkawi 2010; McWhorter 2007) are unlikely, as the speakers surrounding Arabic-speaking cities were likely proficiently bilingual, with a large number of children acquiring the Arabic natively, particularly slaves and children of slaves. I also propose a principle I call the “Bedouin Paradox,” that Bedouin dialects often appear more conservative than city dialects because they are less likely to participate in linguistic changes originating in cities, but are also more likely to represent a quite recent linguistic stratum in a given area, challenging the notion that Bedouin dialects reflect a more ancient stratum of the Arabic language than settled dialects.

Finally, I use samples of the demonstratives and pronouns from seventy modern dialects to illustrate my approach. I reconstruct the pre-Islamic speech communities that were distinguished by variations in these features, and I show how these speech communities were transformed by migration into new contexts. By linking the linguistic and social data, I am able to argue that the primary source of Arabic

speakers in the expansions hailed from the southern Hijaz and Yemen, and that these speech communities are the primary ancestors of modern Arabic dialects, though we are able to reconstruct some aspects of the linguistic behavior of dialects from the eastern Arabian peninsula.

References:

Labov, William. 2001. *Principles of Linguistic Change, Volume 2: Social Factors*. Blackwell.

LePage, Robert, and Andrée Tabouret-Keller. 1985. *Acts of Identity: Creole-based approaches to language and ethnicity*. Cambridge: Cambridge University Press.

McWhorter, J. 2007. *Language Interrupted: Signs of Non-Native Acquisition in Standard Language Grammars*. Oxford scholarship online. Oxford University Press, USA. [http://books.google.com/books?id=iEff\\_YCrGpgC](http://books.google.com/books?id=iEff_YCrGpgC).

Milroy, Lesley. 2008. *Social Networks*. *Handbook of Language Variation and Change*, ed. by J. K. Chambers, Peter Trudgill, and Natalie Schilling-Estes. Wiley-Blackwell.

al Sharkawi, Muhammad. 2010. *The Ecology of Arabic: A study of Arabicization*. Brill.

Versteegh, Kees. 1984. *Pidginization and Creolization: The Case of Arabic*. Benjamins.

Zelinsky, Wilbur. 1992. *The Cultural Geography of the United States: A Revised Edition*. Englewood Cliffs, NJ: Prentice Hall.

Session: Friday, 5:30 PM

**Grammaticalization of the Motion Verb 'rah' as a  
Prospective Aspect Marker in Syrian Arabic**

*Najib Jarad, University of Sharjah*

This paper attempts to show that the future uses of the Syrian Arabic *rah* are said to have grammaticalized from the Classical/Standard Arabic lexical verb *rāḥ* 'go', to the extent that it can be analyzed as a prospective future marker for which there is an intention. The path that this motion verb follows is from andative (1) through purposive (2) to future intention uses (3, 4):

- |    |   |                   |
|----|---|-------------------|
| 1. | (ʔana) <b>rayih</b> ʕal -dʒ-dʒamʕa<br>I going to -the-university<br>'I am going to university'  | [lexical verb]    |
| 2. | (ʔana) <b>rayih</b> šūf rifʔāt-i hnīk<br>I FUT- see friends-my there<br>'I am going to see my friends there'  | [future particle] |
| 3. | (ʔana) <b>rah</b> rūh ʕas-sīnəmə baʕdēn<br>I FUT-go to-the-cinema afterwards<br>'I am going to go to the movies afterwards'                           | [future particle] |
| 4. | l-madrəsə <b>ha-təʕlīn</b> n-nətāyidʒ bukra<br>The-school FUT-announce the-results tomorrow<br>'The school is going to announce the results tomorrow' | [prefix]          |

The grammaticalization path from a lexical verb *rāḥ* 'go' to a prospective future particle in Syrian Arabic resulted in changes at the semantic, syntactic, and phonological levels. First, the transition from lexical to grammatical meaning is seen in contexts where the particle *rah* co-occurs with the lexical verb *rāḥ*. The second stage involves extension in semantic content. As its meaning extended from movement to intention and then to prediction, *rāḥ* started to appear with inanimate subjects which lack volition, as in *rah tuʔaʕ l-kase* 'The glass is going to fall'.

The grammaticalization of the lexical verb *rāḥ* also resulted in dessemanticization or semantic bleaching. In Classical/Standard Arabic, the verb *rāḥ* is a two-place predicate which subcategorizes for an external NP and an internal NP <agent, patient> and/or an external NP and a PP <theme, PP goal>. The <agent, patient> structure is never used in Syrian Arabic. Therefore, the argument structure for the verb *rāḥ* in Syrian Arabic is <theme, PP goal>. When *rāḥ* is reanalyzed as a prospective future marker (*rah*), it is not compatible with any argument structure. In other words, the theta-roles of *rāḥ* are bleached out leaving only the prospective relation of futurity. Furthermore, grammaticalization is a process which affects the phonology of the grammaticalizing item. Thus, prior to grammaticalization, there existed a word boundary, which was reduced through grammaticalization to a morpheme boundary, and finally to an indivisible fused form (*ha-*). In generative terms, this means that grammaticalization involves upward reanalysis; i.e., the loss of movement steps (Roberts & Roussou, 2003). The basic structure assumed in this paper involves an AspP projection placed between TP and vP:

[<sub>TP</sub> T [<sub>AspP</sub> Asp [<sub>vP</sub> v [<sub>VP</sub> V]]]]

In this structure, the lexical verb is merged in V and then moved to v and T. V in the lower VP assigns the GOAL theta-role while the small v assigns the THEME theta-role. The reanalysis of *rāḥ* as a prospective future marker means that this marker is directly merged in Asp, i.e. reanalyzed as an element of Asp. The main motivation for this structure comes from the fact that the past tense auxiliary *kān*, which is merged in T, consistently precedes the prospective marker *rah*.

***Jiim* and the Class of Sun Letters: A Historical and Dialectological Perspective**

Aaron Freeman, University of Pennsylvania

This study investigates historical interaction between the Arabic post-alveolar obstruent *ǧ* (*jiim*, Standard [dʒ]) and the rule of *l*-Assimilation, by which *l* of the definite article assimilates to a following coronal consonant (Ex. 1). We identify three distinct variants of the rule with respect to treatment of *ǧ*. From distributional and historical evidence, we argue that two of the variants have coexisted since the beginning of the historical period, while the third is a recent development in (Cairene) Egyptian Arabic driven by the velarization of *ǧ* (cf. Hary 1996, Blanc 1969). Our sources are grammars and texts of Standard Arabic and the modern spoken dialects (Bergman 2004, Cowell 1964, Erwin 2004, Harrell 2004, Holes 2005, Ingham 1994, Wright 1964, among others).

The behavior of *ǧ* is key to our analysis. This phoneme has a wide variety of pronunciations, all coronal except for Egyptian [g] (see table). It derives from proto-Semitic \*g (Watson 2002), and eighth-century descriptions of Arabic record three distinct pronunciations: a prestige form [j] and non-normative [ʒ] and [g] (Blanc 1969). Cowan (1960) and Kaye (1972) reconstruct ‘proto-Colloquial Arabic’ \*ʒ from the spoken dialects, assuming a diglossic situation in proto-Arabic which has been challenged by Versteegh (1984) and others.

Among Arabic varieties with coronal *ǧ*, Standard Arabic and many Peninsular Arabian dialects exclude *ǧ* from the set of coronals triggering *l*-Assimilation (Grammar I below), while all others generally include *ǧ* in this class of sounds (Grammar II). In Cairene Arabic, non-coronal *ǧ* optionally triggers *l*-Assimilation along with [k] (Grammar III).

<i>Grammar</i>	<i>Phonetic Value of ǧ</i>	<i>Assimilatory Environment</i>	<i>Dialects</i>
I	Coronal [dʒ, ʒ, j, dʒ]	All coronals except ǧ	Standard; Peninsular Arabian
II	Coronal [ʒ, dʒ, ʒ, j]	All coronals	North African; Levantine; Iraqi
III	Non-Coronal [g]	All coronals (obligatory); velar stops (optional)	Cairene Egyptian

Given the presence of Grammar II across the majority of dialect groups, the archaic character of Grammar I, and the early attestation of variable *ǧ*, we propose an early Arabic alternation between Grammar I, with phonologically dorsal *ǧ*, and Grammar II with coronal *ǧ*. The nature of this early Arabic variation is suggested to be originally diatopic in nature, but may have included a diglossic dimension.

We interpret Grammar III as a later development from Grammar II, triggered by rephonemicization of *ǧ* to /g/ in Cairene Arabic. We see this as phonological evidence against the continuity of /g/ in the majority Cairene urban dialect (cf. Woidich and Zack 2009), and support for reintroduction of /g/ into Cairene from an external source. Influenced by this change, *l*-Assimilation split into separate coronal and velar rules, as evidenced by the differential obligatoriness of the two environments and the extension of the velar environment to include /k/. The absence of similar developments in an Omani dialect with velar *ǧ* (Brockett 1985), lying outside the Grammar II area, lends support to this analysis for Cairo.

Finally, we consider the phonological description of Grammar I, positing a palatal stop /j/, rather than velar /g/, as the optimal underlying representation of *ǧ* for most Arabian dialects. This representation both reduces opacity for dialects with phonetically palatal *ǧ* and preserves surface contrasts for *q* in dialects with /g/ < *q*. We further argue that velar patterning of *ǧ* with respect to root co-occurrence restrictions (McCarthy 1994) is a historical rather than a synchronic phonological phenomenon, and is not sufficient evidence for a velar representation of *ǧ* in Grammar I.

Ex. 1	(a)	/al-ʃams/	→	[aʃ-ʃams]	‘the sun’	(Standard Arabic)
	(b)	/al-qamar/	→	[al-qamar]	‘the moon’	

Session: Friday, 6:30 PM

**The Evolution of *Dād***

*Salman Al-Ani and Anthony Woodhams, Indiana University*

The sound *Ḍād* is the least used of the consonants in the Arabic language, but it is also one of the most controversial. Regardless of this controversy, Arabs pride themselves on calling their language “The Language of *Ḍād*”. The original description of *Ḍād* as given by the famous grammarian Sibawayh is that of a voiced lateralized dental fricative. This is quite different from the modern phonetic descriptions of the consonant, but there are a few locations in Yemen which still pronounce *Ḍād* with lateralization, similar to the way in which Sibawayh described. This provides us with evidence that Sibawayh’s description was indeed correct. Today, the realization of *Ḍād* is usually described in one of two ways: either as a voiced emphatic dental stop, or as a voiced emphatic interdental fricative, albeit sometimes with variation.

The *Ḍād* phenomenon is thus fascinating in its description and distribution. For example, in Iraq the vast majority of Iraqis pronounce the *Ḍād* as a fricative and not as a stop. Thus, in Iraq, the most Eastern of the Arabic-speaking countries, there is no distinction between the consonant *Ḍād* and the voiced emphatic interdental fricative *ād*. On the other hand, in Morocco, the most Western country in the Arabic-speaking world, exactly the opposite situation is found; there is no distinction between *Ḍād* and *ād*, and *Ḍād* is used in place of both consonants in speech. The purpose of this study is to: (1) examine how the original *Ḍād* consonant and its description evolved from how Sibawayh documented it, and (2) the historical context surrounding the distribution of the *Ḍād* across the Arabic-speaking world. In order to independently and scientifically analyze *Ḍād*, spectrograms of modern realizations were collected and examined. From our research, new light can be shed on how exactly the Arabic language, specifically its emphatic consonants, interacted with the languages that Arabic came into contact with. Furthermore, we can explore in depth how the *Ḍād* was realized after the introduction of Arabic and provide new theories as to the reasons for the new realizations that arose.

**Session: Saturday, 8:30 AM**

**Compensatory Lengthening: Evidence from Child Arabic**

*Eman Abdoh, King Abdulaziz University*

This paper examines the phonological phenomenon of compensatory lengthening (henceforth CL), wherein the loss of an element of a representation (i.e. a consonant or vowel) triggers a CL of another neighboring phonetic element. It examines this phenomenon in child Arabic phonology, and in particular in child Hijazi Arabic, a dialect spoken in Jeddah, Makkah, and Madina in Saudi Arabia. It investigates this phonological process in early word acquisition within the framework of the Prosodic Theory (McCarthy & Prince 1986, 1990) and Moraic Theory (Hayes, 1985, 1989). Cross-sectional spontaneous data were collected from twenty two monolingual children (aged from 1;0 to 1;9), living in Jeddah, Saudi Arabia, by recording their speech using the object-naming technique in near natural settings and analyzed using a qualitative approach.

The study aims to test if children, who acquire Arabic, a moraic language sensitive to syllable weight and characterized by vowel-length contrast, may use the CL strategy in case of coda consonant or vowel deletion and whether they follow a universal path in this respect (Bernhardt & Stemberger, 1998; Kehoe, 2002; Fikkert, 1994, and Ota, 1998, 2003; Song, J. & K. Demuth, 2008) or influenced by their language-specific phonology. Adopting the moraic conservation approach (Hayes, 1989), the study investigates and provides a mora-based analysis of two types of CL: V- lengthening and C- lengthening. Factors such as adjacency, directionality, and perceived similarity that play a role in motivating CL are also considered.

The analysis results have shown that the subjects' early word productions differ from the target/ adult forms, exhibiting both V- lengthening (e.g. /liʕba/ → [li:ba] 'toy') and C- lengthening (e.g. /dabdu:b/ → [daddu:b] 'teddy bear'), that were used to compensate for a deleted phonetic segment both in mono- and disyllabic forms. Monophthongization of diphthongs accompanied by V-lengthening is evident in the data too (e.g. /laimu:n/ → [li:mu:n] 'lemon'). The moraic analysis has provided a satisfying explanation of CL in Arabic (i.e. lengthening fills a mora that has been vacated as the result of a deletion rule) and it accounts for both V-lengthening and C- lengthening by arguing that the trigger is always a moraic segment. As to C-lengthening, the results also show a preference for left-to-right directionality (i.e. the trigger precedes the target) (e.g. /ʔarnab/ → [ʔannab] 'rabbit'), but there are few cases of right-to-left directionality [ʃukran] → [sukkan] 'thank you'). Trigger and target adjacency is satisfied in both cases. The study concludes that Arabic-speaking children go through a similar universal path in using CL, but despite this, it emphasizes the importance of investigating the impact of the ambient language and the role of language specific phonologies.

**Session: Saturday, 9:00 AM**

**Linguistic Transfer in Learning English as a Second Language in Typically Developing Arabic Heritage School-Age Speakers**

*Reem Khamis-Dakwar, Adelphi University, and Elabbas Benmamoun, University of Illinois at Urbana-Champaign*

Based on the 2000 census, Arab Americans comprise today 0.42% of the population in the United States (U.S. Bureau of the Census, 2005), and the number has likely increased in the last decade. Arabic language has been rated among the top ten languages among English Language Learners (LLEs) in the US (Batalova & Margie, 2010). Many second generation Arab Americans are referred to as heritage language learners and speakers of Arabic. Heritage language learners are usually exposed to Arabic as their native language in early childhood, however this exposure to and use of Arabic lessens or ceases later, particularly in the teen and college years (Albirini, Benmamoun, & Saadah, 2011), a pattern that is typical of other heritage languages in the US such as Chinese, Russian and Spanish.

Research on Arabic language development as a native language is sparse and resources used for language assessment in the Arab world are mainly translated versions of English resources (Omar 1973, Elgibali, 2003; Khamis-Dakwar & Crowley, 2005; Patel & Khamis-Dakwar, 2005; Wiig & El-Halees, 2000; Yaakobi, Hadie, & Khamis-Dakwar, 2003, Al-Jenaie 2010; Khamis-Dakwar & Froud, 2012 ). Similarly, there is rare literature on successive and simultaneous Arabic-English bilingual development in heritage speakers. These studies can contribute to our theoretical understanding of second language acquisition and issues of language maintenance and loss and may be a valuable resource for language assessment and treatment of Arabic-English bilinguals.

The present study used word structure elicitation task, sentence repetition, and sentence formulation to examine use of English phonology and morphosyntax in a group of 40 normally developing 6-9 year olds typically developing Arab-American children from New York and New Jersey, who have two Arab parents speaking Arabic at home, and found to be English dominant. Children's parents country of origin varied (8 Syrian, 7 Egyptian, 5 Jordanian, 5 Palestinian and either Lebanese or Egyptian, 6 to Palestinian parents, 2 had Moroccan parents, 2 Syrian and Yemeni parents, 1 Iraqi, 1 Algerian, 1 Tunisian, and 1 Yemeni family). The study examined the effect of L1 (Arabic), linguistic domain, and structural differences (such as tense and agreement markers), and age on children's English productions.

The findings revealed that children's speech productions (phonologically) were less affected by their first language background, while transfer effects of L1 were more evident in their morphosyntactic productions. More specifically, children performed better on items requiring use of tense markers than agreement markers (i.e. Arabic speaking children with typical language development early learners of English as a second language had more errors producing regular and irregular plural as well as possessive than in producing past tense, future tense, and present progressive

markers at age 6-7 years). Moreover, children had particularly poor performance in producing English irregular plural, derivational morphemes (derivational nouns, derivational adjective, and superlative) and pronouns (possessive pronoun, subjective pronoun, reflexive pronoun, and objective pronoun) at 6-8 age groups. These findings complement our understanding of the process of language learning in heritage speakers in light of the switch from Arabic dominance to English dominance (Albirini et al, 2012). We discuss the results with respect to the multiple factors that affect second language learning in heritage speakers in terms of linguistic transfer effect of the dominant language and home culture, age, structural differences, as well as the levels of exposure and immersion to the two languages with development. Together, these findings point to future directions for the study of language development in Arabic English bilinguals to better understand the interlanguage continuum in heritage speakers in light of the shift from L1 dominant to L2 dominance through development.

Albirini, A. & Benmamoun, E. (2012). Aspects of second language transfer in the oral production of Egyptian and Palestinian heritage speakers. *International Journal of Bilingualism*.

Aljenaie, K. (2010). Verbal inflection in the acquisition of Kuwaiti Arabic. *Journal of Child Language*, 37(4), 841-863.

Batalova, J., & Margie, McHugh (2010). *Top Languages Spoken by English Language Learners Nationally and By State*. Washington, DC: Migration Policy Institute.

Elgibali, A. (2003). *Lu at al-Tifl al- arabi*. Cairo: al-Khanjy Press.

Khamis-Dakwar, R., & Crowley, C. (November, 2005). Language testing for Arabic speaking individuals in Israel. *American Speech & Hearing Association Convention*, San Diego.

Khamis-Dakwar, R., & Froud, K. (2012). Aphasia, Language, and Culture: Arabs in the US. In: Gitterman, M., Goral, M., & Obler, L. (Eds.). *Multilingual aphasia book (275-291)*. Bristol: Multilingual Matters.

Omar, M.K. (1973). *The acquisition of Egyptian Arabic as a native language*. The Hague: Mouton.

Patel, R., & Khamis-Dakwar, R. (2005). An AAC training program for special education teachers: A case study of Palestinian Arab teachers in Israel. *Journal of Augmentative and Alternative Communication*, 21(3), 205-217.

U.S. Bureau of the Census (2005). *We the people of Arab ancestry in the United States*. Census 2000 special reports. Washington, DC: U.S. Government Printing Office.

Wiig, E.H., & El-Halees, Y. (2000). Developing a language screening test for Arabic-speaking children. *Folia Phoniatica et Logopaedica*, 52, 260-274.

Wilson, M.E. (1996). Arabic speakers language and culture, here and abroad. *Topics in Language Disorders*, 16, 4, 65-80.

Yaakobi, G., Hadie, A., & Khamis-Dakwar, R. (2003). A preliminary study on the effect of phonological awareness on acquiring basic skills (reading and writing) in Arabic. *Alrisala*, 13, 1-17.

**Session: Saturday, 9:30 AM**

**Can L2 Primes Trigger L1 Translation Targets in Masked Priming?  
New Evidence from a highly proficient English-Arabic Bilingual**  
*Mahmoud Azaz and Kenneth Forster, University of Arizona*

Since its introduction by Forster and Davis (1984), the masked priming technique has created new psycholinguistic pathways to explore the organization of the bilingual mental lexicon. Many psycholinguistic studies that used this technique concluded that there is an asymmetrical pattern in lexical decision tasks in cross-language priming studies: priming goes only in one direction from first language (L1) to second language (L2) and not from L2 to L1 even at high levels of proficiency (e.g. de Groot & Nas, 1991 for Dutch-English bilinguals; Sanchez-Casa et al. 1992 for Spanish-English; Gollan et al. 1997, for Hebrew-English; Jiang, 1999 for Chinese – English, and Grainger & Frenck-Mestre, 1998, for French-English.) This asymmetrical pattern has remained a puzzle and it has been explained in terms of the variable strength of lexical routes (Kroll and Stewart, 1994) and the conceptual interconnections of the L1-L2 representations (Keatley et al. 1994).

This study hypothesizes that the task demands may be one of the key variables that may explain this asymmetrical pattern. Using the case study approach in the current treatment, one late high proficiency bilingual was instructed to make decisions whether L1 targets are translation equivalents for L2 primes. Since the participant was not conscious of the prime, getting priming effect would mean that L2 primes are able to activate some semantic representation in the L2-L1 direction once task demands are changed.

Over 12 sessions in 6 days, the participant was tested on 72 high frequent translation equivalent pairs used in the L2-L1 direction. The subject was forced to make decisions between two alternatives as either related or unrelated. Arabic primes appeared for 50 milliseconds (ms) followed by the English targets. The related and unrelated conditions were counterbalanced. The participant completed the first session in 10-15 minutes and then the second session. All items were presented on a laptop-controlled display using the DMDX software. He was instructed to give his responses by pressing the YES key if the English word (the target) is a translation equivalent to Arabic (the prime) and press the NO key if it is not.

Using the mixed design effect on Reaction times for correct responses, there was a significant main effect for priming (4.904),  $p = 0.0001$ , error rate is significant,  $p = 0.0001$ , and the interaction between the two is significant,  $p = 0.0352$ . The average success rate was 61.6 %, which was above chance. This result shows an effect for the L2 primes in triggering the L1 targets. The 50 ms duration between the prime and the target was sufficient to activate semantic information that enabled the participant to make his decision about the relatedness of the prime and the target.

Moreover, task demands overrode script differences as Arabic and English use different writing systems.

The paper concludes that the nature of the task is a key factor that needs to be considered when making conclusions about the participants' performance in the asymmetrical priming pattern. As in a semantic discrimination task, when subjects are asked to respond based on the masked word, they exert more effort to discern it. Results of the previous work and this study taken together suggest that priming seems to trigger two distinct processes according to the task at hand. The lexical decision task requires a certain type of activation different from the translation task that requires the subject to go back and forth to the prime to make the right decision regarding whether the prime is related to the target or not. In addition, it could be also possible that semantic activation might be a necessary condition for semantic priming, but not a sufficient condition for lexical decision.

**Session: Saturday, 10:00 AM**

**Gender and Plausibility in the Disambiguation of Relative Clauses in L2 Arabic**

*Abdelaadim Bidaoui, Rebecca Foote, and Mahmoud Abu Nasser, University of Illinois at Urbana-Champaign*

The nature of sentence processing mechanisms in second language (L2) learners has been the topic of a number of recent studies. The debate about the functioning of these mechanisms has centered on whether L2 learners who have acquired their second language after puberty make use of the same types of information as native speakers do when processing sentences in the L2. According to the Shallow Structure Hypothesis (SSH; Clahsen & Felser, 2006), the syntactic representations constructed by late L2 learners “are shallower and less detailed than those of native speakers” (32). Because of this, learners are forced to rely on non-structural rather than structural cues to meaning when parsing the L2. The goal of the current study is to test this hypothesis within the context of relative clause (RC) disambiguation in L2 Arabic.

Relative clause disambiguation has been examined in several L2 sentence processing studies. These studies have looked at ambiguity resolution in sentences such as (1). (1) Someone shot [<sub>NP1</sub>the servant] of [<sub>NP2</sub>the actress] who was on the balcony.

The RC ‘who was on the balcony’ can be attached either to the first noun phrase (NP1) or to the second noun phrase (NP2). If the RC is attached to NP2, ‘the actress’, it is referred to as a case of low attachment. If the RC is attached to NP1, ‘the servant’, it is referred to as a case of high attachment. Native speakers vary cross-linguistically in terms of which NP they prefer to attach RCs to in the absence of disambiguating cues. However, L2 learners do not pattern with native speakers of the L2 in their RC attachment preferences. Instead, they either seem to make use of universal parsing strategies to decide which NP to attach to in the L2 (Dussias, 2003), or they do not show a clear attachment preference (Papadopoulou & Clahsen, 2003). Based in part on this finding, Clahsen and Felser (2006) make a clear distinction between the processing strategies of adult L2 learners and those of native speakers. They argue that L2 learners form a less-detailed structure than native speakers do while processing an L2, and furthermore, that they show reliance on lexical information to parse sentences rather than structural information. To test the validity of the SSH, the present study examines RC attachment and disambiguation in Arabic and asks the following research questions: (1) What type of RC attachment do native speakers and learners of Arabic favor, high or low? (2) What type of cues are used for RC ambiguity resolution in Arabic, structural, non-structural, or both, and does this differ in native speakers and L2 learners? In this study, gender is manipulated to test usage of structural cues, whereas plausibility is manipulated to test usage of non-structural cues. Based on the predictions of the SSH, native speakers of Arabic should make use of both structural and non-structural information in RC disambiguation, that is, both gender and plausibility

cues. L2 learners, on the other hand, should only be able to make use of plausibility cues. Gender should not play a role in their disambiguation of RCs in Arabic.

Ten native speakers of Arabic and ten L1 English learners of Arabic as an L2 participated in one online task (self-paced reading) and two offline tasks: a RC attachment preference task and a grammaticality judgment task (used to verify knowledge of gender in the L2 participants). The self-paced reading task included 48 experimental sentences with RCs that could be attached to either of two NPs, along with 96 fillers. Half of the experimental sentences contained gender and half contained plausibility cues that indicated which of the two NPs the RC modified. It was predicted that reading times at the region in the sentence where the cues appeared would indicate participants' attachment preferences in the following way: when the particular cue present in the sentence disambiguated toward the preferred NP, reading times would be faster than when the cue disambiguated toward the non-preferred NP. Analyses of the results of the online task indicated that both groups relied on plausibility cues ( $p=.024$ ) rather than gender cues ( $p=.692$ ), although the GJT results indicated that the L2 learners were aware of gender information in sentences containing RCs. In the preference task, both groups showed a preference for HA. Based on these results, the SSH was partially confirmed. As expected, L2 learners relied on plausibility cues and did not transfer their L1 attachment preference to English. However, in contrast to the claims of the SSH that native speakers make use of both structural and non-structural information in sentence parsing in this experiment, they did not rely on gender information in the disambiguation of RC; they only relied on plausibility. Implications for theories of L1 and L2 sentence processing will be discussed.

## Session: Saturday, 10:45 AM

10:45-11:15     **Verb Inflections in Adolescents with Down Syndrome: Experimental approach**  
*Khawla Aljenaie, Kuwait University*

Several studies on young individuals with Down Syndrome (DS) have shown that expressive language skills present particular challenges and generally are more impaired than receptive skills (Caselli et al., 1998; Chapman, Hesketh, & Kistler, 2002; Laws & Bishop, 2003). Crosslinguistics experimental studies show that the implementation of verb inflections is affected (Ring & Clahsen 2005; Stathopoulou & Clahsen 2010; Eadie, Fey, Douglas, & Parsons, 2002; Hesketh & Chapman, 1998). Besides, individuals with DS continue to produce shorter and less complex utterances than Typically Developing (TD) children of the same nonverbal mental age as they get older (Caselli et al., 2008; Chapman et al., 1998; Price et al., 2008; Rosin et al., 1988).

The current study is the first to investigate verb morphology in Arabic speaking individuals with DS. The aim of this research is to assess the implementation of the verb inflections in Kuwait Arabic-speaking adolescents with DS compared with TD children. Two groups of participants were elicited: ten Kuwaiti Arabic speaking adolescents with DS (aged 13.9–16.5) and 20 TD children (4;0-5;6) whose chronological age matched with the mental age of the DS group. The participants were tested in two tasks. First, they were presented with a wordless picture book and were asked to tell a story. Then in an elicited production procedure they were presented with 20 pictured stimuli (15 targeted verbs as well as 5 distractors) that show gender distinction and were asked to complete the utterance. For example, *hathi elbint tesoog ow hatha el-walad.....* ‘this girl is riding, and this boy..’.

The results of the storytelling showed that VPs were significantly less frequent in adolescents with DS than TD children ( $p < .000$ ). As for the types of verb inflections, the results show that the DS used significantly more imperfective verb inflections than the perfective ones ( $p < .003$ ). Moreover, the DS participants produced shorter and less complex structures than the TD children.

In the elicitation task, there was a significant difference between the two groups in terms of the correct use (DS 63% and TD 85%, respectively). The results also showed that in 12% and 4% of the time, respectively, the DS participant and TD children left out the required imperfective verb inflection in obligatory contexts, and that difference was significant. When the two groups omitted the required verb inflection, a behavior which was found in both tasks, they used the imperfective bare stem in place of the imperfective verb inflection. For example, they would complete the sentence *hathi elbint te-soog ow hatha el-walad .....* ‘this girl is riding, and this boy ...’ saying *soog* instead of *ye-soog* ‘he rides’. This result supports the Aljenaie’s (2010) finding, and Benmamoun’s (1999, 2000) argument that the imperfective bare stem is the default form of Arabic language. Overall, the results revealed that in spite of the rare use of verb inflections by the DS participants in

storytelling, they inflected the verbs appropriately when prompted. This emphasizes the importance of utilizing a variety of language samples.

By exploring the acquisition of Arabic morpho-syntactic structures, the current research provides information on typical and impaired acquisition that would have clinical implications. For future research, further experimental and clinical observations using larger samples and different populations will assist in generalizing the current results to other Arabic speakers with language impairment.

#### Selected references

Aljenaie, K. (2010). Verbal inflection in the acquisition of Kuwaiti Arabic. *Journal of Child Language*, **37** (4), 841–863.

Benmamoun, E. (1999). Arabic morphology: the central role of the imperfective. *Lingua* **108**, 175–201.

Benmamoun, E. (2000). *The feature structure of functional categories: a comparative study of Arabic dialects*. New York: Oxford University Press.

Hesketh, L. J. & Chapman, R S. (1998). Verb use with individuals with Down Syndrome. *American Journal on Mental Retardation*, **103**: 288-304.

Stathopoulou, N. & H. Clahsen (2010). The perfective past tense in Greek adolescents with Down Syndrome. *Clinical Linguistics & Phonetics* 24: 870-882.

Ring, M. & H. Clahsen (2005). Morphosyntax in Down's Syndrome: Is the Extended Optional Infinitive Hypothesis an option? *Stem-, Spraak- en Taalpathologie* **13**: 3-13.

**Session: Saturday, 11:15 AM**

**Agrammatism in Moroccan Arabic-Speaking Subjects**

*Samir Diouny, Chouaib Doukkali University, Morocco*

Individuals with agrammatism show selective impairments in functional categories. On the one hand, some researchers (Friedmann & Grodzinsky, 1997) have attributed many linguistic difficulties experienced by agrammatic aphasics to a deficit in the hierarchy of functional projections. Other theories (Kolk, Kok, Zevenbergen, & Haverkort, 2007), however, suggest that agrammatic production results from limitations in processing capacities (Kolk, 1995). The purpose of this study was to investigate the use of functional categories by two Moroccan-speaking agrammatic aphasics across different tasks. The functional categories studied are: tense affixes, subject-verb agreement and noun and adjective agreement. The availability of verbal affixes and grammatical features was investigated using spontaneous speech, recitation of Koranic verses, picture description and picture-matching tasks. In task 1 the participants were asked to talk about their family life, work, history of illness, hobbies and preferences for entertainment. In task 2, the participants were asked to describe the “cookie theft” picture (Kaplan & Goodglass, 1983). In task 3, they had to recite Koranic verses. In task 4, the participants were asked to name 60 pictures (30 depicting actions, and 30 depicting an object). The participants achieved high correctness scores for object naming, while scores for action naming were below average. They also achieved high correctness scores for verbal and nominal agreement, while tense affixes were impaired. To account for the observed disassociations, we suggest that the production deficit in Moroccan Arabic agrammatism cannot be explained in terms of a structural account, but rather in terms of a processing account that takes the view that inaccessible syntactic knowledge affects the subjects’ ability to produce verbal affixes.

**Conclusion:** The study provides further evidence that tense production is selectively impaired in agrammatic aphasia. On this account, we posited a processing account to explain the problems experienced by individuals with agrammatism.

**References**

- Akkal, A. (1993) *Syntactic derivation of the inflections of the verb in Moroccan Arabic*. Doctorat d'Etat. Faculté des Lettres I, Casablanca.
- Akkal, A. & Gonegai, A. (2000) On the status of agrS in null subject languages. In J. Lecarme, J. Lowenstamer, & U. Shlonsky (Eds.), *Research in Afro-Asiatic Grammar*, 202, 1-22.
- Avrutin, S. (2000). Comprehension of discourse-linked and non-discourse linked questions by children and Broca’s aphasics. In Y. Grodzinsky, L. Shapiro, & D. Swinney (Eds.), *Language & the brain*. San Diego: Academic Press.
- Baker, M. (1985). The mirror principle and morphosyntactic explanation. *Linguistic Inquiry*, 16, 373-416.
- Beeke, S., Maxim, J., & Wilkinson, R. (2008). Rethinking agrammatism: factors affecting the form of language elicited via clinical test procedures. *Clinical Linguistics and Phonetics*, 22 - 323.

- Benedet, M. J., Christiansen, J. A., & Goodglass, H. (1998). A crosslinguistic study of grammatical morphology in Spanish-speaking and English-speaking agrammatic patients. *Cortex*, 34, 309–336.
- Benmamoun, A. (2000). *The feature structure of functional categories: A comparative study of Arabic dialects*. New York: Oxford University Press.
- Burchert, F., Swoboda-Moll, M., & de Bleser, R. (2005). Tense and agreement dissociations in German agrammatic speakers: underspecification vs. hierarchy. *Brain & Language*, 94, 188–199
- Chomsky, N. (2000). Minimalist inquiries: the framework. In R. Martin, D. Michaels, & J. Uriagereka (Eds.), *Step by step* (pp. 89–155). Cambridge, MA: MIT Press.
- Clahsen, H. 2008. Chomskyan syntactic theory and language disorders. In: M. J. Ball, M. Perkins, N. Mueller & S. Howard (eds.), *The Handbook of Clinical Linguistics*. Blackwell: Oxford, pp.165-183.
- Diouny (2005). Neurolinguistics: Some Aspects of Moroccan Arabic Agrammatism. Unpublished P.h.D. Thesis. Hassan II University, Faculty of Letters and Human sciences, Mohammedia, Morocco
- El Alaoui Faris. M., Ettahiri, L. Fennich, R. & Benbelaid, F. (in press). *The Boston diagnostic aphasia examination: Arabic version*. Faculty of Medicine, Rabat.
- Faroqi-Shah, Y., & Thompson, C. K. (2004). Semantic, lexical, and phonological influences on the production of verb inflections in agrammatic aphasia. *Brain and Language*, 89, 484–498.
- Faroqi-Shah, Y., & Dickey, M. W. (2009). On-line processing of tense and temporality in agrammatic aphasia. *Brain and Language*, 108, 97-111.
- Fassi-Fehri, A. (1993). *Issues in the structure of Arabic clauses and word order*. London: Kluwer Academic.
- Friedmann, N. (2000). Moving verbs in agrammatic production. In R. Bastiaanse & Y. Friedmann, N. (2001). Agrammatism and the psychological reality of the syntactic tree. *Journal of Psycholinguistic Research*, 30, 71–90.
- Friedmann, N. (2006). Speech production in Broca’s agrammatic aphasia: Syntactic Tree Pruning. In Y. Grodzinsky & K. Amunts (Eds.), *Broca’s Region (2006)*. Oxford University Press.
- Friedmann, N., & Grodzinsky, Y. (1997). Tense and agreement in agrammatic production: pruning the syntactic tree. *Brain & Language*, 56, 397–425.
- Jarema, G. (1998). The breakdown of morphology in aphasia: A cross-language perspective. In B. Stemmer & H. A. Whitaker (Eds.), *Handbook of neurolinguistics (pp. 221-234)*. New York: Academic Press
- Kolk, H. (1998). Disorders of syntax in aphasia: Linguistic-descriptive and processing approaches. In B. Stemmer & H. A. Whitaker (Eds.), *Handbook of neurolinguistics*. (pp. 249-260). New York: Academic Press.
- Lee, J., Milman, L. H., & Thompson, C. K. (2008). Functional category production in English agrammatism. *Aphasiology*, 22, 893-905.
- Menn, L. & Obler, L. (Ed.), (1990). *Agrammatic aphasia: A cross-language narrative sourcebook*. John Benjamins.
- Mimouni, Z. & Jarema, G. (1997). Agrammatic aphasia in Arabic. *Aphasiology*, 11, 125-144
- Stavrakaki, S. & Kouvava, S. (2003). Functional categories in agrammatism: Evidence from Greek. *Brain & Language*, 86, 129-141.
- Wenzlaff, M., & Clahsen, I. (2004). Tense and agreement in German agrammatism. *Brain & Language*, 89, 57–68
- Yarbay Duman, Y., & Bastiaanse, R. (2009). Time reference through verb inflection in Turkish agrammatic aphasia. *Brain and Language*, 108, 30-39.

Session: Saturday, 11:45 AM

**An Exploration of the Formation of Canonical Agreement Morphosyntactic Features in English-Arabic Interlanguage**

*Boshra El-Ghazoly, Indiana University*

The present research investigates the acquisition of canonical agreement features in the English-Arabic interlanguage system. Successful acquisition of (Modern Standard) Arabic canonical agreement would entail the addition of a new *u*Feature, *u*Gender. The Failed Functional Features Hypothesis (FFFH) (Hawkins & Chan, 1997) predicts English-Arabic learners' failure in acquiring Arabic canonical agreement morphosyntactic properties as it involves the acquisition of *u*Gender, a feature absent from English. Contrary to FFFH, White, Valenzuela, Kozłowska-Macgregor, and Leung (2004), Bruhn De Gravito & White (2002), and Tanner (2007) demonstrate the acquirability of *u*Gender by L2ers whose L1s lack the feature beyond puberty. In support of Full Access stands (FA), White et al. (2004) attribute incompleteness and variability in learners' interlanguage to reasons other than the inaccessibility of UG featural inventory beyond puberty. Four groups of participants took part in the study, a native speaker-group (N=15), three learner-groups at 3 levels of proficiency in Arabic L2 (Intermediate (N=15), Upper-intermediate (N=15), and Advanced (N=15)). The data were collected using two tasks: a Grammaticality Judgment Task and a Picture Description task. The results were discussed in light of two main models of L2 acquisition, the Failed Functional Features Hypothesis (FFFH) (Hawkins & Chan, 1997), and Full Transfer/ Full Access (FTFA) (Schwartz & Sprouse 1994, 1996). The results demonstrate that English-Arabic learners were able to add *u*Gender features successfully: Learners showed high accuracy rates in SVO and Noun- adjective constructions requiring canonical agreement.

References

Bruhn de Gravito, J., & White, L. (2002). L2 acquisition of Spanish DPs: The status grammatical features. In A. T. Perez-Leroux & J. Liceras (Eds.), *The acquisition of Spanish morphosyntax: the L1/L2 connection* (pp.153-178). Dordrecht, Netherlands: Kluwer.

Hawkins, R., & Chan, Y-C. (1997). The partial availability of universal grammar in second language acquisition: The failed functional features hypothesis. *Second Language Research*, 13, 187-226.

Schwartz, B. D., & Sprouse, R. A. (1994). Word order and nominative case in nonnative language acquisition: A longitudinal study of (L1 Turkish) German interlanguage. In T. Hoekstra & B. D. Schwartz (Eds.), *Language Acquisition Studies in Generative Grammar* (pp. 317-368). Amsterdam: John Benjamins.

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the “full transfer/full access” model. *Second language Research*, 12, 40-72.

Tanner, D. (2007). *Agreeing that agreement isn't the problem*. In M. Bowles, R. Foote, S. Perpiñán & R. Bhatt (Eds.), *Selected Proceedings of the 2007 Second Language Research Forum* (pp. 22-34). Somerville, MA: Cascadilla Press.

White, L., Valenzuela, E., Kozłowska-Macgregor, M., & Leung, Y-K. I. (2004). Gender agreement in nonnative Spanish: Evidence against failed features. *Applied Psycholinguistics*, 25, 105-133.

**Session: Saturday, 12:45 PM**

**Social Media Arabic**  
*Muhammad Abdul Mageed, Indiana University*

### **Introduction**

In spite of the importance of social media in the ongoing Arab Spring, there is not much literature on Arabic computer-mediated discourse (CMD). In the current paper, we report attempts to partially bridge this gap by investigating the use of Arabic in five social media genres: (1) YouTube comments, (2) chat turns, (3) Twitter's tweets, (4) Wikipedia Talk Pages (WTP), and (5) Web for a (WF). Previous studies on Arabic social media (e.g., Authors [2011a]; Authors (2011b); Palfreyman & Khalil [2003]) are limited in both scope and context, and hence their results are far from generalizable.

### **Goal and Research Questions**

Our goal in the current work is to analyze the linguistic features, identify language variety employed (i.e., Modern Standard Arabic [MSA], the modern standard variety of the language, vs. dialectal Arabic [DA]), and sentiment expressed in social media Arabic (SMA). We thus aim to answer the following specific research questions:

- RQ1: What are the linguistic features characteristic of SMA?
- RQ2: What is the language variety distribution in SMA genres?
- RQ3: To what extent is SMA 'opinionated' (i.e., carries, e.g., positive and negative sentiment)?

### **Data Collection, Sampling, and Methods**

For the study, we collected five different data sets, each corresponding to one genre, as follows:

1. A comments corpus of 1200 Arabic comments crawled from 60 randomly sampled videos from the YouTube channels of Alarabiya (<http://www.youtube.com/user/alarabiya>) and Aljazeera (<http://www.youtube.com/user/aljazeerachannel>) (20 comments on each of 30 videos from each of the two channels).
2. A turns corpus of more than 3000 hours of chat data collected from 'Egypt's room' in Maktoob chat, the chat system of the popular Arabic portal Maktoob, which was acquired by Yahoo! in 2009. The bigger data set is composed of 384 chat sessions collected between December 2008 and February 2010 and totals 58,985,108 million words participated by 5,052,206 unique users.
3. A Twitter corpus of 233,309 tweets (henceforth the Twitter Arabic Corpus [TAC]) automatically extracted from a two-month Twitter stream between November 11, 2009 and February 1, 2010 as reported in Petrovic et al. (2010).
4. A WTP corpus of 30 highly interactive WTPs pertaining various topics (e.g., political and religious).
5. A WF corpus of about 30,000 threaded conversations from a variety of popular Arabic WFs.

We employ corpus linguistics (McEnery & Wilson, 2001) methods to study the linguistic features in the five data sets. For both sentiment and language variety, we employ content analysis (Bauer, 2000). For the manual content analysis, we employ a grounded-theory approach to analyze a random sub-corpus of the 3181 YouTube comments, 2798 chat turns from 10 randomly sampled sessions, 3015 tweets extracted from TAC, 3008 sentences randomly sampled from WTP, and 3097 sentences randomly sampled from WF. Two college-educated native speakers of Arabic were trained to label the data for both sentiment and language variety. The labeling was performed independently, and then annotators met and resolved the differences.

The unit of analysis is the comment, tweet, chat turn, or sentence, as appropriate. For sentiment, each annotator assigned one of 4 labels: (1) Objective (OBJ), (2) Subjective-Positive (S-POS), (3) Subjective-Negative (S-NEG), and (4) Subjective-Mixed (S-MIXED). For language variety, each annotator assigned one of 2 possible tags: (1) MSA and (2) dialect (DA).

## **Results**

Results of the analysis of linguistic features show that whereas Arabic YouTube comments, tweets, WTP and WF sentences do not differ strikingly from 'offline' Arabic, the chat turns share more of the features characteristic of some varieties of online language (e.g., paralinguistic and prosodic features). Regarding the analysis of language variety, for YouTube, 50.56% of comments are in MSA, while the rest are in dialects. For the chat data, 95% is in Egyptian Arabic and only 5% is in MSA. For Twitter, it was found that 51.37.1% of the tweets are in MSA, whereas the rest are dialectal. Both the WTP and the WF data sets had 97% MSA, while the rest in each case is in DA. For the sentiment analysis, for YouTube, it is found that 50.50% of the comments are negative, 21.65% are positive, 34.67% are mixed, and 11.72% are objective. For chat data, 33.16 % of the comments are negative, 58.86% are positive, 34.76% are mixed, and 11.72% are objective. Regarding Twitter, it was found that 25.17% of the comments are negative, 16.01% are positive, 11.44% are mixed, and 47.36% are objective. As to WTP, 33.71% of the comments found to be negative, 21.68% positive, 4.52% mixed, 40.09% are objective. For WF data, it was found that 33.16% of the comments are negative, 35.60% are positive, 12.69% are mixed, and 18.60% are objective.

## **Discussion and Conclusion**

At a first glance, the finding that in both tweets and YouTube comments no specific linguistic features (e.g., clippings and abbreviations) are employed may seem surprising. However, this is expectable as in Arabic the practice of dropping vowels results in a level of ambiguity and employing abbreviations would heighten such ambiguity and renders text unintelligible. In addition, since in chat turns more of these features (e.g., letter repetition) are used, it seems that the synchronic (or lack thereof) nature of a genre is one of the primary deciding factors as to the employment of linguistic features characteristic of digital English, for instance. The synchronic nature of chat also imposes a vast majority of dialectness, which

contrasts with more MSA use in the other asynchronous genres (i.e., YouTube comments, tweets, WTP, and WF). The finding that more positive than negative chat turns and WF sentences are posted shows that these two genres are different from e.g., listserves (Thompson & Ahn, 1992) where users are less likely to know one another. The political topics of comments on the YouTube channels and WTPs also seems to be a factor that imposes more disagreement and hence negativity. The more negative use of Twitter is also indicative of the employment of this medium for argumentative causes (e.g., ones characteristic of the Arab Spring). In conclusion, we note that SMA is different from other digital forms of, e.g., Indo-European language (e.g., English) not only because of cultural differences but because of the linguistic features of the language itself (e.g., its orthographic underspecification) as well.

## References

- Bauer, M. (2000). Classical content analysis. In M. Bauer & G. Gaskell (Eds.), *Qualitative researching with text, image and sound* (pp. 131-151). Thousand Oaks, CA: Sage.
- McEnery, T., & Wilson, A. (2001). *Corpus linguistics* (Vol. 1). Edinburgh.
- Petrovic, S., Osborne, M., & Lavrenko, V. (2010). The Edinburgh Twitter Corpus. *Computational Linguistics in a World of Social Media*, 25.
- Palfreyman, D. and Khalil, M. (2003). A Funky Language for Teenzz to Use?: Representing Gulf Arabic in Instant Messaging. *Journal of Computer-Mediated Communication*, 9(1).
- P.A. Thompson and D.K. Ahn, 'To Be or Not To Be: An Exploration of E-Prime, Copula Deletion and Flaming in Electronic Mail.', *Et Cetera: A Review of General Semantics*, 49, 146-164, (1992).

Session: Saturday, 1:15 PM

Variation in the Representation of Arabic Consonants in Facebook  
Duaa Abu Elhija Mahajna, Indiana University

This paper analyzes variation in Electronic Ameiyas, written versions of colloquial Arabic used by young Arabic speakers in electronic media such as Facebook (Palfreyman & Al-Khalil, 2003; Garra, 2007). In recent years, electronic writing has begun to replace print writing, and as in the case of the development of the printing press in the 15th century, this new technology has also influenced the written colloquial forms of languages, as young people all over the world are formulating and devising new conventions for writing their indigenous spoken languages electronically. This linguistic development is now occurring in many languages in the world, particularly ones characterized by diglossia, such as Arabic, Persian, and languages of the Indian subcontinent like Tamil, and Bengali.

This paper presents the results of an analysis of variation in the consonantal system in the Electronic Ameiyas written on Facebook of several countries in the Arab World (Kuwait, United Arab Emirates (UAE), Jordan, Lebanon, Palestine/Israel, Egypt, and Morocco). The purpose of the study is to examine how the Latin writing system is used to represent the Arabic consonantal system amongst different users in different countries and to see to what extent standardization is occurring. I focused on two matters in the consonantal system: letters which do not have equivalents in the Latin writing system, such as *ḥaaʔ* /h/, and *sʕaad* /sʕ/; and consonants pronunciations that are subject to dialectal variation, such as *qaf* /q/.

The data was gathered from about 170 Facebook friends from the seven Arab countries mentioned above. There were 95 males and 75 females, whose ages are from 17 to 30. The results of the data show that there is not a complete standardization yet; however, young people are writing more similarly to each other within general dialect areas. This is because they begin to accept a standard regional way to write. For instance, in some of the villages in Palestine/Israel where *qaf* is pronounced as /g/ and *kaf* as /tʃ/, young people write these two characters as <2> /ʕ/ and <k> /k/ respectively instead of <g> and <ch>; thus, following the prestigious dialect of the city rather than the local form of these villages. Moreover, Palestinian women in Jordan who pronounce *qaf* as /g/ also tend to write <2> /ʕ/ which is the standard form amongst Palestinians in Jordan particularly women.

Moreover, I found that in Kuwait and UAE the letter *jeem* is written as <j> or <y> since *qaf* is sometimes pronounced [g] and written as <g>. The letter *chaf* /tʃ/ in the final position is written as <j>. In this area the <'> (apostrophe) precedes the number, as in <'3> for *ghain*, which is the primary form in this area. Finally, EA users in this area stress pharyngealized phonemes as in Classical Arabic, using numerals, <'9> or <'6> for *Dhaaʕ*, <9> for *sʕaad*, and <6> for *tʕaaʕ*.

Lebanese youngsters as well as Egyptians, rarely express pharyngealizations, and when they do they use capitalization, e.g. *s<sup>ʕ</sup>aad* is <S> or <s>. They use <2> for *qaf*, <t>/<s> for *thaa'*, and <d>/<z> for *dhaal*. The form <ch> for *sheen* is common and this is most likely a French influence and is particularly associated with Christians. The use of the digraph <gh> is the most common form for *ghain*. The letter *qaf* in Lebanon, Palestine/Israel and Egypt has the standard written form of <2> (glottal stop /<sup>ʕ</sup>/) on Facebook.

Gender differences are especially apparent in Jordan. This is because females adopt the more "feminine way of writing" with less pharyngealizations while males use pharyngealizations a lot. For example, girls tend to write the letter *t<sup>ʕ</sup>aa'* as <t> while boys use <6>. Jordanians most often use <2> for *qaf*. Sometimes <8> or <g> is used for *qaf* by Palestinian males, Jordanian males and females.

Moroccans tend to use digraphs more than any other group, they use <kh> for *khaa'*, <gh> for *ghain*, and <ch> for *sheen*. Moreover, the letter *qaf* is written as <9>. Consequently, if we examine all these different variants, there seem to be three dialectal areas in terms of Electronic Ameiya, based upon the representation of the most common sound *qaf*:

- (1) Kuwait/UAE (the <8>/<g> area)
- (2) Egypt/Israel-Palestine/Lebanon (the <2> area)
- (3) Morocco (the <9> area).

The dialect in Jordan is basically a mixture of group (1) and group (2) (the <2> and <8>/<g> areas).

### References:

Garra, I. (2007) M.A Thesis. *From a Dialect into a Language: The case of English and Arabic*. University of Haifa, Israel.

Palfreyman, D. & Al- Khalil, M. (2003). 'A Funky Language for Teenzz to use: Representing Gulf Arabic in instant Messaging '. In *Journal of Computer Communicated Communication*.9 (1).

## Session: Saturday, 1:45 PM

### Arabic Chat Alphabet: A data-oriented analysis of variation in Latinized Arabic

Paul Rodrigues, C. Anton Rytting, and Timothy Buckwalter, University of Maryland

Arabizi is a non-standard and non-deterministic Arabic chat alphabet that uses Latin characters to replace Arabic script. Character choices may be either phonologically similar to their replaced sound, or visually similar to a form of their replaced Arabic script character. Entire messages can be written in Arabizi. [3] found that Egyptian college graduates used Arabizi for 67% of informal e-mails, 61% of chat text and 13% of formal e-mails. Similar results were found for UAE students in [2].

Arabizi has regional variation stemming from the large phonetic differences between Arabic dialects, and the lack of standardization causes additional individual variation within the regional variants. Several papers have published tables of correspondence observed regionally between the Arabizi and Arabic characters [2, 4], but these are inadequate for natural language processing (NLP). No work has published tables statistically induced from data, which would allow for accurate understanding of character preferences, character substitution weights suitable for computational transliteration, and computational region identification from text. This paper will introduce a parallel Arabizi/Arabic script data set, provide a linguistic analysis of this dataset, and an explanation on how this data-set can be used to build Arabizi-aware NLP applications.

Amazon Mechanical Turk was used to crowd-source an aligned corpus of Arabizi and Arabic. Users were presented with Arabic text, and asked to retype it in Arabizi. All prompts were in Arabic script, and precaution was taken to prevent the use of existing Arabizi/Arabic script conversion software. The submitted samples were hand aligned to the word, with one side of the alignment corresponding to a single token. If connected morphemes in the Arabic stimuli were separated by a space character in the Arabizi sample, the Arabic side of the alignment would have one token, and the Arabizi side may have several. Conversely, if two Arabic words were conjoined in the Arabizi, several tokens in the Arabic script aligned with one single token in the Arabic. Clear spelling errors and misalignments were removed from the data. An example alignment of nine words from a subset of speakers is presented in Table 1. This phrase-aligned word list then served as input to the Many-2-Many Aligner, a statistical alignment system designed for character-to-character alignment [1]. This system was configured to associate one Arabic script character to one Arabizi character, and to maximize the joint character-to-character probability. Arabizi characters were allowed to be associated with a null character (indicated by an underscore). A subset of the character alignments are presented in Table 2.

In the full paper, we discuss the collected data, highlighting the personal and the regional differences. We discuss the computational usage of the computed statistical tables, such as computational machine transliteration and regional identification.

User ID	ولما	كان	من	الضروري	أن	يتولى	القانون	حماية	حقوق
1	wa lama	kana	mina	addarouriyi	an	yatawala	al9anouno	7imayata	7o9o9i
2	walama	kana	mina	darori	an	yatawala	9anon	7imayat	7o9o9
...	...	...	...	...	...	...	...	...	...
12	walamma	kan	men	aldaroori	an	yatawalla	al qanoon	7emayat	(SKIPPED)

Table 1: Intraspeaker Arabizi Word Alignment, SAMPLE of Users for Abstract Submission

Arabizi	Arabic Script with Probability
7	( ح 93.376 ) ( _ 6.624 )
d	( د 72.576 ) ( ض 22.664 ) ( ذ 2.309 ) ( ظ 1.066 ) ( _ 0.795 ) ( ل 0.591 )
2	( ا 39.862 ) ( ا 19.798 ) ( ء 15.168 ) ( ئ 10.112 ) ( ا 8.834 ) ( ا 1.618 ) ( و 1.336 ) ( ق 1.240 ) ( _ 1.176 ) ( ة 0.857 )
j	( ج 100.000 )
...	...

Table 2: Arabizi to Arabic Alignment, SAMPLE of Alphabet for Abstract Submission, with Weights

## References

- [1] S. Jiampojarn, G. Kondrak, and T. Sherif. Applying many-to-many alignments and hidden Markov models to letter-to-phoneme conversion. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, pages 372–379, Rochester, New York, April 2007. Association for Computational Linguistics.
- [2] D. Palfreyman and M. al Khalil. “A funky language for teenzz to use”: Representing Gulf Arabic in instant messaging. *Journal of Computer Mediated Communication*, 9(1), 2003.
- [3] M. Warschauer, G. R. E. Said, and A. Zohry. Language choice online: Globalization and identity in Egypt. *Journal of Computer Mediated Communication*, 7(4), 2002.
- [4] M. A. Yaghan. “Arabizi”: A contemporary style of Arabic slang. *Design Issues*, 24(2):39–52, 2008.