

Workshop on Arabic Corpus Linguistics
Lancaster University, April 11-12 2011

Multi-factorial methods for exploring contextual factors in the usage of Modern Standard Arabic COME verbs

Dana Abdulrahim, John Newman, and Sally Rice
University of Alberta



COME verbs in MSA

- The phenomenon
 - Four lexical items denoting basic COMING event in MSA أتى، جاء، قدم، حضر
 - **'ata** أتى, **ǧa'a** جاء, **qadima** قدم, and **ḥaḍara** حضر
 - Interchangeable in some contexts but not in others

COME verbs in MSA

My grandmother **came** to the airport to meet me

جدتي إلى المطار لتستقبلني

أتت

'ata

جاءت

ġa'a

قدمت

qadima

حضرت

ḥaḍara

come:3SGF:PERF jaddati ila al maṭar litastaqbilani

COME verbs in MSA

My grandmother did not **come** to the airport to meet me

جدتي إلى المطار لتستقبلي

{
تأت 'ata
تجئ ? ġa'a
تقدم ? qadima
تحضر ḥaḍara
لم

lam **come:3SGF:JUSS** jaddati ila al maṭar litastaqbilani

COME verbs in MSA

Their desire to see her **came** after they had heard about her success

رغبتهم في رؤيتها بعد سماعهم بخبر نجاحها

أتت

'ata

جاءت

ġa'a

قدمت

* qadima

حضرت

* ḥaḍara

come:3SGF:PERF raġbatuhum fi ru'yatiha ba'da
sama'ihim bixabar naġaḥiha

COME in bilingual dictionaries

- English-MSA
 - **Al Muhit Oxford Dictionary (2008)**
 - أتى 'ata, جاء ġa'a, حضر ḥaḍara
 - وصل waṣala, أتى 'ata, قدم qadima
 - **Concise Oxford Dictionary (1982)**
 - جاء ġa'a
 - حدث ḥadaṭa
 - حصل ḥaṣala
 - نتج nataġa

Theoretical approach

- Constructionist framework
 - The behaviour of a lexical item is best understood in its **context of usage** and not in isolation
 - The need to scrutinize the entire **lexico-syntactic frame** in which the lexical item under question appears, including
 - syntactic structures in which it appears
 - morphological inflections associated with it
 - other lexical items it occurs with

Methodological approach

- Corpus-based
 - Corpora provide naturally-occurring, contextualized uses of the individual lexical item under study
 - Corpora provide a large amount of linguistic data allowing for a quantitative treatment of the investigated phenomenon

Methodological approach

- Quantitative analysis
 - Constructing a data frame for every lexical item under investigation
 - Gries (2006), Gries & Divjak (2009), Gries and Otani (2010), among others
 - Annotating a large number of concordance lines for an extensive set of linguistic features (morphological, syntactic, semantic)
 - Apply multi-factorial statistical tests to this data frame
 - Cluster analysis
 - Hierarchical Configural Frequency Analysis

Corpus query

- arabiCorpus.byu.edu
- Downloaded all hits of finite verb forms of **'ata**, **ğā'a**, **qadima** and **ḥaḍara** from
 - newspapers
 - modern literature
 - non fiction
- 500 random concordance lines per verb

Constructing the data frame

- 500 random concordance lines per verb were coded for over 40 factors including the following
 - **Verb morphology**
 - Tense, aspect, morphological aspect/mood
 - **Syntactic features of the construction**
 - Syntactic transitivity, negation, complements, adverbials, etc.
 - **Sentential subject category**
 - Individual, event, activity, notion, state, time, etc.
 - **Spatial setting**
 - Inclusion of goal, source, manner, setting, temporal, purposive, etc. phrases

Constructing the data frame: 'ata

Ate DATA				GENRE	VERB
sort word	ten words after	word	ten words before	GENRE	VERB
كما	«اقسم بالله العلي العظيم ان اكون مخلصا للوطن والشعب وان	يأتي:	وكان نص اليمين كما	NEWS	ATE
ما	«ان رابين وبيريز سوف يقتلان. وسترين ذلك بعينيك. اننا نندخل	يأتي:	احد المتطرفين اليهود. وقد جاء على لسان هذا المتطرف ما	NEWS	ATE
الأرياني	«في اطار مواصلة المباحثات الثنائية الهادفة الى ترسيم ما تبقى	تأتي	وأشارت «وكالة الأنباء السعودية» الى أن زيارة الأرياني	NEWS	ATE
	«كلاروس» في مستويي تجهيز مقتعين جدا:	تأتي		NEWS	ATE
فجأة	« ولأنه «يسقط علينا».	يأتي	«الصوت الذي يخلق الناس»، وعلى «انه كينونة مفارقة: لأنه فجأة	NEWS	ATE
الاجواء	احيانا بثمار يانعة طبقا للقاعدة الشرعية التي تنص على ان	تأتي	لعبت الاجهزة الاعلامية دورا في تأجيحها، الا ان تلك الاجواء	NEWS	ATE
	استقالة كوباس قبل أيام من تصويت مجلس الشيوخ في باراجواي	وتأتي		NEWS	ATE
	استكمالاً لخطوات سابقة ، ولقياس حجم الرد الفلسطيني بالدرجة الاولى،	ويأتي	تندرج في اطار مخطط معن ومعلوم على المستوى الصهيوني ،	NEWS	ATE
للبحرين	استكمالاً للجولة التي بدأها قبل ايام على دول المجلس للاجتماع	تأتي	آل خليفة وقال قبل عودته الى الرياض ان زيارته للبحرين	NEWS	ATE
لدمشق	استكمالاً للقاء القمة الثنائية التي انعقدت في اللاذقية لبلورة موقف	تأتي	صعيد آخر ، قال مصدر وزاري ان زيارة الحريري وبويز لدمشق	NEWS	ATE
ولم	الا لصد المعتدي عليكم، واذا يبدو من احد معتمديها شيء	تأت	حضرة الامير ليس للحكومة طمع فيك ولا في بلدك، ولم	NEWS	ATE
وقد	الاجوبة على النحو الآتي:	أنت	السؤال على الشباب العربي في البلدان التي شملتها العينة. وقد	NEWS	ATE
ان	الاختراقات للانحة الانتلافية، محدودة باثنين او ثلاثة على الاكثر. وقد	تأتي	لكن مصادر بقاعية توقعت ان	NEWS	ATE
	الاعتداء عقب يوم واحد من هجوم آخر علي الحورة العلمية	ويأتي		NEWS	ATE
	الاعلان السعودي عن الصفقة الجديدة نتيجة محادثات طويلة تمت في	ويأتي		NEWS	ATE
	الاعلان عن القبض على الايرانيين الثمانية من دون الاقصادح عن	ويأتي		NEWS	ATE
وقد	الايام ليقل ان الفاعلين الاساسيين ليسوا عربا ولا مسلمين.	تأتي	هذه هي بعض الانفعالات وردود الغضب على المسلمين والعرب وقد	NEWS	ATE
وعندما	الإستلام	يأتي	البداية ، لا أن يكتتب المواطن على السيارة بسعر وعندما	NEWS	ATE
الكبيرة	الإصلاح الجماعي عندما يكون العمل الفردي محدودا وقاصرا.	ويأتي	شوارعنا وبيوتنا وإدارتنا ومساجدنا، والبدء من الأمور الصغيرة ثم الكبيرة،	NEWS	ATE
و	الإعانات الرئيسية من المتبرعين خاصة في شهر رمضان بالإضافة إلى	تأتي	الأجهزة الصوتية المخصصة للأفراح والمؤتمرات وريعتها يعود للدار مباشرة، و	NEWS	ATE
أيام	الأمر الذي لا يدفع، فغادرنا قبل وداعه، فهل أقل من	أناه	الخميس قبل وفاته لقد كان في زيارة للطبيب. وبعد أيام	NEWS	ATE
وعندما	البابا يوحنا بولس الثاني الى لبنان، سيقول بالتأكيد للبنانيين كم	يأتي	وعندما	NEWS	ATE
	البالات في صورة رئيسية من المانيا والنمسا وسويسرا والولايات المتحدة،	تأتي		NEWS	ATE

Constructing the data frame: 'ata

VERB MORPHOLOGY				SYNTACTIC TRANSITIVITY						SUBJECT			
PREFIX	TENSE	ASPECT	PH_ASP/MTRANSITIV	OBJ_TYPE	J_PRON_NJ	FJ_PRON_GJ	PRON_CSUBJ_NUM	SUBJ_PERS	SUBJ_GEN	SUBJ_CAT			
NILL	PRES	SIMPLE	IMPF	NO	NILL	NILL	NILL	NILL	NILL	SING	3RD	MASC	COMMUNI
NILL	PRES	SIMPLE	IMPF	NO	NILL	NILL	NILL	NILL	NILL	SING	3RD	MASC	content
NILL	PRES	SIMPLE	IMPF	NO	NILL	NILL	NILL	NILL	NILL	SING	3RD	FEM	EVENT
NILL	PRES	SIMPLE	IMPF	NO	NILL	NILL	NILL	NILL	NILL	SING	3RD	FEM	OBJECT
NILL	PRES	HAB	IMPF	NO	NILL	NILL	NILL	NILL	NILL	SING	3RD	MASC	NOTION
NILL	PRES	HAB	IMPF	NO	NILL	NILL	NILL	NILL	NILL	SING	3RD	FEM	STATE
NILL	PRES	SIMPLE	IMPF	NO	NILL	NILL	NILL	NILL	NILL	SING	3RD	FEM	ACTIVITY
NILL	PRES	SIMPLE	IMPF	NO	NILL	NILL	NILL	NILL	NILL	SING	3RD	MASC	GROUP
NILL	PRES	SIMPLE	IMPF	NO	NILL	NILL	NILL	NILL	NILL	SING	3RD	FEM	EVENT
NILL	PRES	SIMPLE	IMPF	NO	NILL	NILL	NILL	NILL	NILL	SING	3RD	FEM	EVENT
NILL	NILL	NILL	JUSS	NO	NILL	NILL	NILL	NILL	NILL	SING	3RD	NILL	GROUP
NILL	PAST	SIMPLE	IMPF	NO	NILL	NILL	NILL	NILL	NILL	SING	3RD	FEM	COMMUNI
NILL	NILL	NILL	SUBJN	NO	NILL	NILL	NILL	NILL	NILL	SING	3RD	FEM	ACTIVITY
NILL	PRES	SIMPLE	IMPF	NO	NILL	NILL	NILL	NILL	NILL	SING	3RD	MASC	ACTIVITY
NILL	PRES	SIMPLE	IMPF	NO	NILL	NILL	NILL	NILL	NILL	SING	3RD	MASC	COMMUNI
NILL	PRES	SIMPLE	IMPF	NO	NILL	NILL	NILL	NILL	NILL	SING	3RD	MASC	COMMUNI
NILL	NILL	NILL	IMPF	NO	NILL	NILL	NILL	NILL	NILL	SING	3RD	FEM	TIME
NILL	PRES	HAB	IMPF	NO	NILL	NILL	NILL	NILL	NILL	SING	3RD	MASC	ACTIVITY
NILL	PRES	HAB	IMPF	NO	NILL	NILL	NILL	NILL	NILL	SING	3RD	MASC	ACTIVITY
NILL	PRES	HAB	IMPF	NO	NILL	NILL	NILL	NILL	NILL	SING	3RD	FEM	ACTIVITY
NILL	PAST	SIMPLE	PERF	NO	NILL	NILL	NILL	NILL	NILL	SING	3RD	MASC	NOTION
NILL	FUT	SIMPLE	IMPF	NO	NILL	NILL	NILL	NILL	NILL	SING	3RD	MASC	INDIVIDUA
NILL	PRES	HAB	IMPF	NO	NILL	NILL	NILL	NILL	NILL	SING	3RD	FEM	OBJECT

Constructing the data frame: 'ata

SYNTAX							SPATIAL SETTING					
INTEROG	NEGATIONS	SV_ORDEF	SVC	PP	LOC_ADV	ADVERBIAL	GOAL	GOAL_CO	GOAL_CA	SOURCE	SOURCE_	SOURCE_
NO	NO	NILL	NO	NO	NO	NO	NO	NILL	NILL	NO	NILL	NILL
NO	NO	NILL	NO	NO	NO	NO	NO	NILL	NILL	NO	NILL	NILL
NO	NO	SV	NO	YES	NO	NO	NO	NILL	NILL	NO	NILL	NILL
NO	NO	VS	NO	YES	NO	NO	YES	PP	STATE	NO	NILL	NILL
NO	NO	NILL	NO	NO	NO	YES	NO	NILL	NILL	NO	NILL	NILL
NO	NO	SV	NO	YES	NO	NO	NO	NILL	NILL	NO	NILL	NILL
NO	NO	VS	NO	NO	YES	NO	NO	NILL	NILL	NO	NILL	NILL
NO	NO	NILL	NO	NO	NO	YES	NO	NILL	NILL	NO	NILL	NILL
NO	NO	NILL	NO	NO	NO	YES	NO	NILL	NILL	NO	NILL	NILL
NO	NO	SV	NO	NO	NO	NO	NO	NILL	NILL	NO	NILL	NILL
NO	YES	NILL	NO	YES	NO	NO	NO	NILL	NILL	NO	NILL	NILL
NO	NO	VS	NO	YES	NO	NO	NO	NILL	NILL	NO	NILL	NILL
NO	NO	VS	NO	NO	NO	YES	NO	NILL	NILL	NO	NILL	NILL
NO	NO	VS	NO	NO	YES	NO	NO	NILL	NILL	NO	NILL	NILL
NO	NO	VS	NO	NO	YES	NO	NO	NILL	NILL	NO	NILL	NILL
NO	NO	NILL	NO	NO	NO	YES	NO	NILL	NILL	NO	NILL	NILL
NO	NO	VS	NO	YES	NO	NO	NO	NILL	NILL	NO	NILL	NILL
NO	NO	VS	NO	YES	NO	NO	NO	NILL	NILL	NO	NILL	NILL
NO	NO	NILL	NO	NO	NO	NO	NO	NILL	NILL	NO	NILL	NILL
NO	NO	VS	NO	NO	YES	NO	NO	NILL	NILL	NO	NILL	NILL
NO	NO	VS	NO	YES	NO	NO	NO	NILL	NILL	YES	PP	INDIVIDUA
NO	NO	VS	NO	NO	YES	NO	NO	NILL	NILL	NO	NILL	NILL
NO	NO	VS	NO	YES	NO	NO	YES	PP	GEO_LOC	NO	NILL	NILL
NO	NO	VS	NO	YES	NO	NO	NO	NILL	NILL	YES	PP	GEO_LOC

Constructing the data frame: 'ata

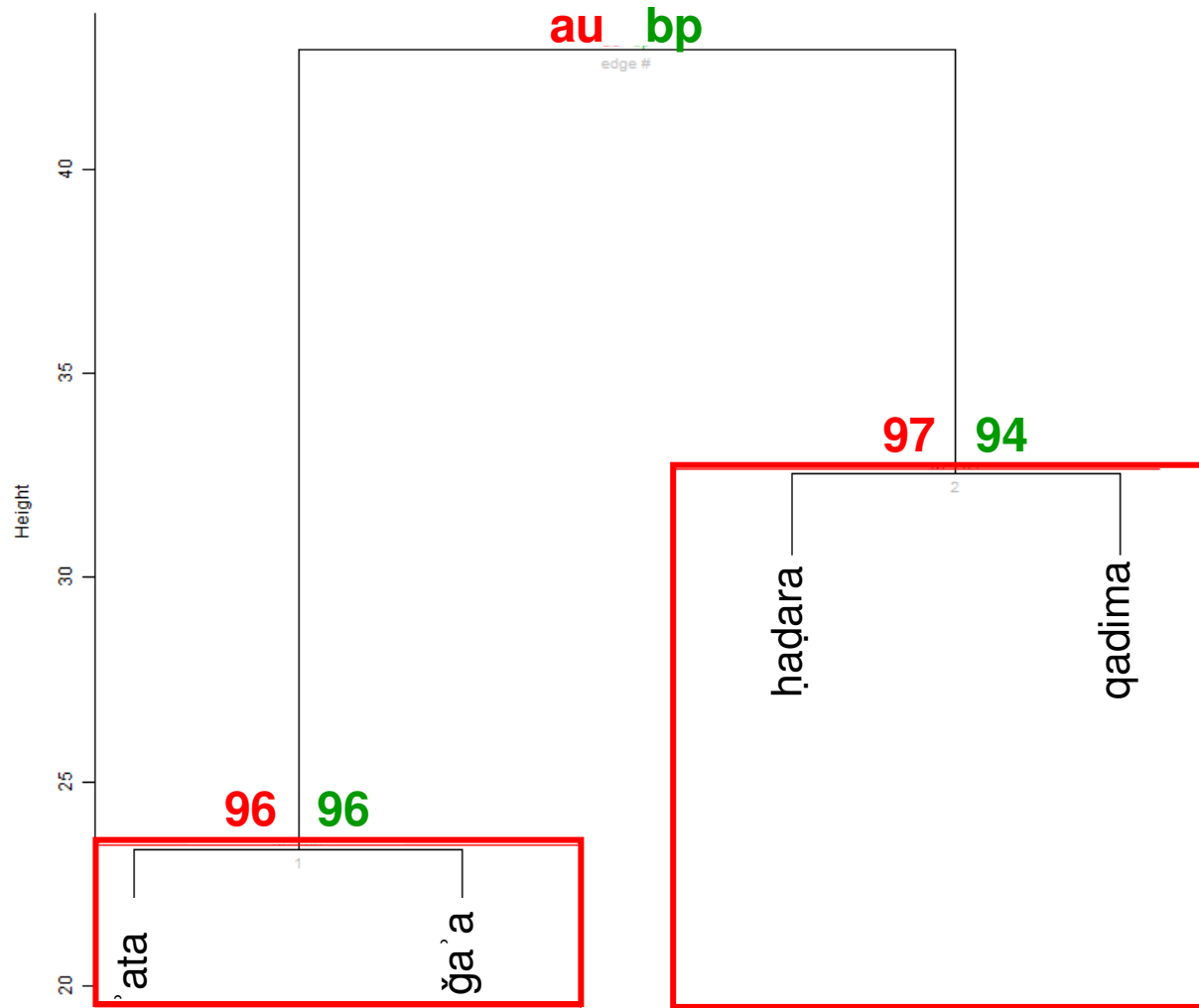
SPATIAL SETTING

MANNER	MANNER_SETTING	SETTING_PATH	PATH_COI	PURPOSI	PURPOSI	COMITATI	COMITATI	TEMPORA	TEMPORA	DEGREE	DEGREE
NO	NILL	NO	NILL	NO	NILL	NO	NILL	NO	NILL	NO	NILL
NO	NILL	NO	NILL	NO	NILL	NO	NILL	NO	NILL	NO	NILL
NO	NILL	YES	PP	NO	NILL	NO	NILL	NO	NILL	NO	NILL
NO	NILL	NO	NILL	NO	NILL	NO	NILL	NO	NILL	NO	NILL
YES	ADVERBIA	NO	NILL	NO	NILL	NO	NILL	NO	NILL	NO	NILL
NO	NILL	NO	NILL	NO	NILL	NO	NILL	YES	PP	NO	NILL
NO	NILL	NO	NILL	NO	NILL	NO	NILL	NO	NILL	YES	LOC_ADV
YES	ADVERBIA	NO	NILL	NO	NILL	NO	NILL	NO	NILL	NO	NILL
YES	ADVERBIA	NO	NILL	NO	NILL	NO	NILL	NO	NILL	NO	NILL
NO	NILL	NO	NILL	NO	NILL	YES	ADVERBIA	NO	NILL	NO	NILL
NO	NILL	NO	NILL	NO	NILL	YES	PP	NO	NILL	NO	NILL
YES	PP	NO	NILL	NO	NILL	NO	NILL	NO	NILL	NO	NILL
YES	ADVERBIA	NO	NILL	NO	NILL	NO	NILL	NO	NILL	NO	NILL
NO	NILL	NO	NILL	NO	NILL	NO	NILL	NO	NILL	YES	LOC_ADV
YES	ADVERBIA	NO	NILL	NO	NILL	NO	NILL	NO	NILL	NO	NILL
YES	PP	NO	NILL	NO	NILL	NO	NILL	NO	NILL	NO	NILL
NO	NILL	NO	NILL	NO	NILL	YES	SVC	NO	NILL	NO	NILL
NO	NILL	NO	NILL	NO	NILL	NO	NILL	NO	NILL	NO	NILL
NO	NILL	NO	NILL	NO	NILL	NO	NILL	NO	NILL	YES	LOC_ADV
NO	NILL	NO	NILL	NO	NILL	NO	NILL	NO	NILL	YES	PP
NO	NILL	NO	NILL	NO	NILL	NO	NILL	NO	NILL	YES	LOC_ADV
NO	NILL	NO	NILL	NO	NILL	NO	NILL	NO	NILL	NO	NILL
YES	PP	NO	NILL	NO	NILL	NO	NILL	NO	NILL	NO	NILL

Multi-factorial statistical tests

- Cluster Analysis (multiscale bootstrap resampling)
 - R script
 - Behavioral Profiles 1.01
 - Stefan Th. Gries (2009)
 - Applies multiple cluster analyses as a means of resampling parts of data and returns a dendrogram that specifies p -values for all possible substructures

Cluster dendrogram with AU/BP values (%)



Distance: canberra
Cluster method: ward

Multi-factorial statistical tests

- Hierarchical Configural Frequency Analysis (HCFA)
 - Von Eye (1990)
 - systematically works through *all* possible pair wise combinations of features to determine which combinations are the most significant, i.e. of greater or lesser frequency than would be expected by chance.
 - R script
 - hcfa 3.2.
 - Stefan Th. Gries (2004)

Syntactic variables

VERB	TRANSITIVIT	NEGATION	PP	LOC_ADV	ADVERBIA	Freq	Exp	Cont.chisq	Obs-exp	P.adj.bin	Dec	Q
ATE	YES	26	107.75	62.0238	<	5.06E-21	***	0.043
ATE	.	YES	.	.	.	34	16	20.25	>	4.44E-04	***	0.009
ATE	.	.	YES	.	.	318	302.75	0.7682	>	1.427407	ns	0.009
ATE	.	.	.	YES	.	69	49.5	7.6818	>	0.036394	*	0.01
ATE	YES	64	55	1.4727	>	9.91E-01	ns	0.005
HDR	YES	380	107.75	687.8892	>	4.70E-100	***	0.144
HDR	.	YES	.	.	.	29	16	10.5625	>	0.016781	*	0.007
HDR	.	.	YES	.	.	95	302.75	142.5601	<	8.95E-49	***	0.122
HDR	.	.	.	YES	.	19	49.5	18.7929	<	4.12E-06	***	0.016
HDR	YES	27	55	14.2545	<	0.000145	***	0.014
JAC	YES	15	107.75	79.8382	<	3.71E-29	***	0.049
JAC	.	YES	.	.	.	0	16	16	<	8.44E-07	***	0.008
JAC	.	.	YES	.	.	322	302.75	1.224	>	0.972414	ns	0.011
JAC	.	.	.	YES	.	64	49.5	4.2475	>	2.03E-01	ns	0.007
JAC	YES	90	55	22.2727	>	5.39E-05	***	0.018
QDM	YES	10	107.75	88.6781	<	6.84E-34	***	0.052
QDM	.	YES	.	.	.	1	16	14.0625	<	1.45E-05	***	0.008
QDM	.	.	YES	.	.	476	302.75	99.1431	>	2.03E-23	***	0.102
QDM	.	.	.	YES	.	46	49.5	0.2475	<	2.718335	ns	0.002
QDM	YES	39	55	4.6545	<	0.10929	ns	0.008

Syntactic variables

VERB	TRANSITIVIT	NEGATION	PP	LOC_ADV	ADVERBIA	Freq	Exp	Cont.chisq	Obs-exp	P.adj.bin	Dec	Q
ATE	YES	26	107.75	62.0238	<	5.06E-21	***	0.043
ATE	.	YES	.	.	.	34	16	20.25	>	4.44E-04	***	0.009
ATE	.	.	YES	.	.	318	302.75	0.7682	>	1.427407	ns	0.009
ATE	.	.	.	YES	.	69	49.5	7.6818	>	0.036394	*	0.01
ATE	YES	64	55	1.4727	>	9.91E-01	ns	0.005
HDR	YES	380	107.75	687.8892	>	4.70E-100	***	0.144
HDR	.	YES	.	.	.	29	16	10.5625	>	0.016781	*	0.007
HDR	.	.	YES	.	.	95	302.75	142.5601	<	8.95E-49	***	0.122
HDR	.	.	.	YES	.	19	49.5	18.7929	<	4.12E-06	***	0.016
HDR	YES	27	55	14.2545	<	0.000145	***	0.014
JAC	YES	15	107.75	79.8382	<	3.71E-29	***	0.049
JAC	.	YES	.	.	.	0	16	16	<	8.44E-07	***	0.008
JAC	.	.	YES	.	.	322	302.75	1.224	>	0.972414	ns	0.011
JAC	.	.	.	YES	.	64	49.5	4.2475	>	2.03E-01	ns	0.007
JAC	YES	90	55	22.2727	>	5.39E-05	***	0.018
QDM	YES	10	107.75	88.6781	<	6.84E-34	***	0.052
QDM	.	YES	.	.	.	1	16	14.0625	<	1.45E-05	***	0.008
QDM	.	.	YES	.	.	476	302.75	99.1431	>	2.03E-23	***	0.102
QDM	.	.	.	YES	.	46	49.5	0.2475	<	2.718335	ns	0.002
QDM	YES	39	55	4.6545	<	0.10929	ns	0.008

Syntactic variables

Coefficient of pronouncedness: the higher the value the more entrenched the configuration

VERB	TRANSITIVIT	NEGATION	PP	LOC_ADV	ADVERBIA	Freq	Exp	Cont.chisc	Obs-exp	P.adj.bin	Dec	Q
ATE	YES	26	107.75	62.0238	<	5.06E-21	***	0.043
ATE	.	YES	.	.	.	34	16	20.25	>	4.44E-04	***	0.009
ATE	.	.	YES	.	.	318	302.75	0.7682	>	1.427407	ns	0.009
ATE	.	.	.	YES	.	69	49.5	7.6818	>	0.036394	*	0.01
ATE	YES	64	55	1.4727	>	9.91E-01	ns	0.005
HDR	YES	380	107.75	687.8892	>	4.70E-100	***	0.144
HDR	.	YES	.	.	.	29	16	10.5625	>	0.016781	*	0.007
HDR	.	.	YES	.	.	95	302.75	142.5601	<	8.95E-49	***	0.122
HDR	.	.	.	YES	.	19	49.5	18.7929	<	4.12E-06	***	0.016
HDR	YES	27	55	14.2545	<	0.000145	***	0.014
JAC	YES	15	107.75	79.8382	<	3.71E-29	***	0.049
JAC	.	YES	.	.	.	0	16	16	<	8.44E-07	***	0.008
JAC	.	.	YES	.	.	322	302.75	1.224	>	0.972414	ns	0.011
JAC	.	.	.	YES	.	64	49.5	4.2475	>	2.03E-01	ns	0.007
JAC	YES	90	55	22.2727	>	5.39E-05	***	0.018
QDM	YES	10	107.75	88.6781	<	6.84E-34	***	0.052
QDM	.	YES	.	.	.	1	16	14.0625	<	1.45E-05	***	0.008
QDM	.	.	YES	.	.	476	302.75	99.1431	>	2.03E-23	***	0.102
QDM	.	.	.	YES	.	46	49.5	0.2475	<	2.718335	ns	0.002
QDM	YES	39	55	4.6545	<	0.10929	ns	0.008

Spatial setting variables

VERB	GOAL	SOURCE	MANNER	SETTING	PURPOSIV	COMITATI	TEMPORA	Freq	Exp	Cont.chisq	Obs-exp	P.adj.bin	Dec	Q
ATE	YES	109	202.5	43.1716	<	3.02E-13	***	0.052
ATE	.	YES	69	90.75	5.2128	<	0.073	ms	0.011
ATE	.	.	YES	84	55.25	14.9604	>	0.001231	**	0.015
ATE	.	.	.	YES	.	.	.	62	58.25	0.2414	>	2.613524	ns	0.002
ATE	YES	.	.	40	66.5	10.5602	<	0.002091	**	0.014
ATE	YES	.	36	22	8.9091	>	0.028567	*	0.007
ATE	YES	81	73	0.8767	>	1.476159	ns	0.004
HDR	YES	419	202.5	231.4679	>	1.94E-45	***	0.12
HDR	.	YES	16	90.75	61.5709	<	8.91E-22	***	0.039
HDR	.	.	YES	22	55.25	20.0102	<	1.89E-06	***	0.017
HDR	.	.	.	YES	.	.	.	20	58.25	25.117	<	3.45E-08	***	0.02
HDR	YES	.	.	25	66.5	25.8985	<	2.55E-08	***	0.021
HDR	YES	.	13	22	3.6818	<	0.217274	ns	0.005
HDR	YES	37	73	17.7534	<	1.42E-05	***	0.019
JAC	YES	56	202.5	105.9864	<	7.49E-36	***	0.082
JAC	.	YES	37	90.75	31.8354	<	4.85E-10	***	0.028
JAC	.	.	YES	84	55.25	14.9604	>	0.001231	**	0.015
JAC	.	.	.	YES	.	.	.	147	58.25	135.22	>	1.56E-22	***	0.046
JAC	YES	.	.	63	66.5	0.1842	<	2.883396	ns	0.002
JAC	YES	.	25	22	0.4091	>	2.299648	ns	0.002
JAC	YES	89	73	3.5068	>	0.28237	ns	0.008
QDM	YES	226	202.5	2.7272	>	0.366253	ns	0.013
QDM	.	YES	241	90.75	248.761	>	8.21E-41	***	0.079
QDM	.	.	YES	31	55.25	10.6437	<	0.001886	**	0.012
QDM	.	.	.	YES	.	.	.	4	58.25	50.5247	<	9.78E-20	***	0.028
QDM	YES	.	.	138	66.5	76.8759	>	2.70E-14	***	0.037
QDM	YES	.	14	22	2.9091	<	0.374538	ns	0.004
QDM	YES	85	73	1.9726	>	0.699554	ns	0.006

Morphological variables

VERB	TENSE	ASPECT	MORPH_ASP	Freq	Exp	Cont.chisq	Obs-exp	P.adj.bin	Dec	Q
ATE	PRES	SIMPLE	IMPF	206	18.8432	1858.9029	>	3.13E-136	***	0.094
ATE	NILL	NILL	SUBJN	40	0.1223	13000.126	>	1.04E-82	***	0.02
ATE	NILL	NILL	IMPF	32	0.8442	1149.8143	>	2.63E-36	***	0.016
ATE	NILL	NILL	JUSS	17	0.0724	3959.9503	>	4.49E-32	***	0.008
ATE	FUT	SIMPLE	IMPF	14	2.6846	47.6923	>	0.000418	***	0.006
ATE	NILL	COND	JUSS	2	0.0113	350.7155	>	0.028259	*	0.001
HDR	PAST	SIMPLE	PERF	339	201.6032	93.6389	>	1.54E-18	***	0.076
HDR	FUT	SIMPLE	IMPF	38	2.6846	464.5576	>	9.18E-28	***	0.018
HDR	PRES	SIMPLE	IMPF	38	18.8432	19.4756	>	0.027761	*	0.01
HDR	NILL	NILL	JUSS	18	0.0724	4441.6379	>	1.79E-34	***	0.009
HDR	PRES	HAB	IMPF	11	1.5722	56.5361	>	0.00038	***	0.005
HDR	NILL	NILL	IMPF	9	0.8442	78.792	>	0.000124	***	0.004
HDR	NILL	IMPR	JUSS	2	9.00E-04	4252.4521	>	0.000198	***	0.001
JAC	PAST	SIMPLE	PERF	484	201.6032	395.569	>	3.43E-71	***	0.157
JAC	NILL	COND	PERF	8	0.3751	154.9988	>	3.08E-06	***	0.004
QDM	PAST	SIMPLE	PERF	467	201.6032	349.3769	>	1.28E-63	***	0.148
QDM	PAST	PERT	PERF	15	2.0721	80.6576	>	2.65E-06	***	0.006
QDM	NILL	COND	PERF	9	0.3751	198.3207	>	1.27E-07	***	0.004

Subject-related variables

VERB	SUBJ_NUM	SUBJ_PER	SUBJ_GEN	SUBJ_CAT	Freq	Exp	Cont.chisq	Obs-exp	P.adj.bin	Dec	Q
ATE	SING	3RD	FEM	ACTIVITY	43	8.2328	146.8227	>	2.13E-14	***	0.017
ATE	SING	3RD	FEM	EVENT	32	2.5951	333.1824	>	1.01E-20	***	0.015
ATE	SING	3RD	FEM	NOTION	33	6.1298	117.7853	>	5.62E-11	***	0.013
ATE	SING	3RD	FEM	COMMUNICATION	24	4.9665	72.9432	>	1.47E-06	***	0.01
ATE	SING	3RD	MASC	NOTION	39	20.4464	16.8359	>	0.349507	ns	0.009
ATE	SING	3RD	MASC	ACTIVITY	41	27.4609	6.6752	>	20.17117	ns	0.007
ATE	SING	3RD	FEM	GROUP	20	5.2797	41.0413	>	0.001698	**	0.007
ATE	SING	3RD	MASC	COMMUNICATION	29	16.5661	9.3324	>	7.703274	ns	0.006
ATE	SING	3RD	FEM	OBJECT	14	1.0738	155.5962	>	2.50E-08	***	0.006
HDR	SING	3RD	MASC	INDIVIDUAL	369	164.6162	253.7584	>	1.61E-44	***	0.111
HDR	SING	3RD	FEM	GROUP	25	5.2797	73.6572	>	9.86E-07	***	0.01
JAC	SING	3RD	MASC	content	92	14.4767	415.1414	>	1.48E-39	***	0.039
JAC	SING	3RD	FEM	ACTIVITY	42	8.2328	138.4981	>	1.14E-13	***	0.017
JAC	SING	3RD	MASC	demonstrative	34	7.1637	100.5325	>	7.43E-10	***	0.013
JAC	SING	3RD	FEM	COMMUNICATION	24	4.9665	72.9432	>	1.47E-06	***	0.01
JAC	SING	3RD	MASC	ACTIVITY	52	27.4609	21.9282	>	0.037878	*	0.012
JAC	SING	3RD	MASC	NOTION	37	20.4464	13.4019	>	1.325838	ns	0.008
JAC	SING	3RD	MASC	COMMUNICATION	32	16.5661	14.3791	>	1.051302	ns	0.008
JAC	SING	3RD	FEM	NOTION	18	6.1298	22.986	>	0.162689	ns	0.006
QDM	PL	3RD	MASC	INDIVIDUAL	238	32.2855	1310.757	>	7.39E-121	***	0.105
QDM	DUAL	3RD	MASC	INDIVIDUAL	19	2.3915	115.3416	>	2.84E-08	***	0.008

Selected variables

- All four verbs
 - Tense
 - Subject number
 - Subject category
 - Goal
 - Source
 - 30+ levels

Prototypical uses

	Tense	Subj_Num	Subj_Cat	Goal	Source
<i>'ata</i>	PRESENT	SINGULAR	NON-HUMAN	NO	NO
<i>ğa'a</i>	PAST	SINGULAR	NON-HUMAN	NO	NO
<i>qadima</i>	PAST	PLURAL	HUMAN	.	YES
<i>hađara</i>	PAST	SINGULAR	HUMAN	YES	NO

Prototypical uses

	Tense	Subj_Num	Subj_Cat	Goal	Source
'ata	PRESENT	SINGULAR	NON-HUMAN	NO	NO
ġa'a	PAST	SINGULAR	NON-HUMAN	NO	NO
qadima	PAST	PLURAL	HUMAN	.	YES
ħađara	PAST	SINGULAR	HUMAN	YES	NO

تأتي المسابقة في إطار العلاقات الثقافية بين مصر واليابان

The competition comes within the frame of the cultural relations between Egypt and Japan

وقال أن زيارته للبحرين **تأتي** استكمالاً للجولة التي بدأها قبل أيام على دول المجلس

And he said that his visit to Bahrain comes as a conclusion to his GCC trip that he had started a few days ago

Prototypical uses

	Tense	Subj_Num	Subj_Cat	Goal	Source
<i>'ata</i>	PRESENT	SINGULAR	NON-HUMAN	NO	NO
<i>ġa'a</i>	PAST	SINGULAR	NON-HUMAN	NO	NO
<i>qadima</i>	PAST	PLURAL	HUMAN	.	YES
<i>ħađara</i>	PAST	SINGULAR	HUMAN	YES	NO

وجاء هذا التوضيح بعد مقال نشرته أول من أمس جريدة “الجمهورية”

This clarification came after an article that was published by “Al-Jumhuriya” newspaper two days ago

جاء ذلك في بيان أصدره المجلس الأعلى إثر اجتماعه أمس

This came in a statement issued by the High Council after its meeting yesterday

Prototypical uses

	Tense	Subj_Num	Subj_Cat	Goal	Source
<i>'ata</i>	PRESENT	SINGULAR	NON-HUMAN	NO	NO
<i>ġa'a</i>	PAST	SINGULAR	NON-HUMAN	NO	NO
<i>qadima</i>	PAST	PLURAL	HUMAN	.	YES
<i>ħaħara</i>	PAST	SINGULAR	HUMAN	YES	NO

موجها التحية للذين **قدموا** من بيروت للمشاركة في هذه المناسبة

Greeting those who had come to Beirut to participate in this event

أفضل الرياضيين في العالم **قدموا** إلى أتلانتا

The best athletes in the world came to Atlanta

Prototypical uses

	Tense	Subj_Num	Subj_Cat	Goal	Source
<i>'ata</i>	PRESENT	SINGULAR	NON-HUMAN	NO	NO
<i>ġa'a</i>	PAST	SINGULAR	NON-HUMAN	NO	NO
<i>qadima</i>	PAST	PLURAL	HUMAN	.	YES
<i>ħađara</i>	PAST	SINGULAR	HUMAN	YES	NO

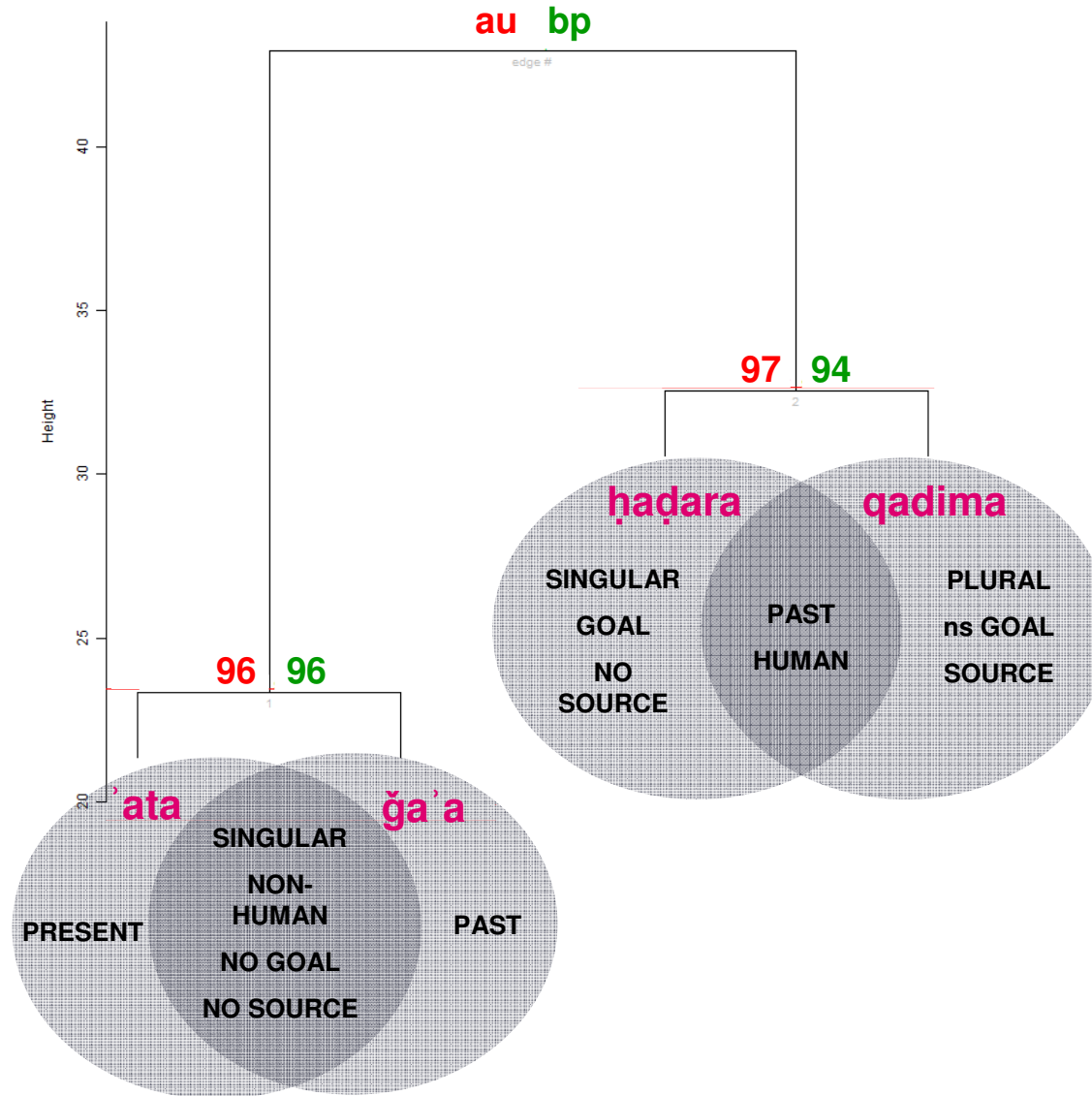
حضر الحفلة سفير لبنان لدى فرنسا ناجي أبي عاصي

Lebanon's ambassador in France, Naji Abi Aasi, attended the party

لقد نظمت اليونسكو مؤتمرا **حضره** أكثر من مئة مفكر ومنتقف عربي

The UNESCO organized a conference, which over a hundred Arab thinkers and intellectuals had attended

Cluster dendrogram with AU/BP values (%)



Conclusion

- The construction of such a multivariate data frame that specifies a wide range of constructional elements is time consuming yet rewarding
- Allows for fine grained synthesis of the data using a wide variety of multi-factorial statistical tests
- Empirical approach to understanding the behavior of lexical items

Thank you!

Dana Abdulrahim

dana.rahim@ualberta.ca

John Newman

john.newman@ualberta.ca

Sally Rice

sally.rice@ualberta.ca

References

Gries, Stefan Th. (2009). *Statistics for linguistics with R: A practical introduction*. Berlin & New York: Mouton de Gruyter.

Gries, Stefan Th. 2006. Corpus-based methods and cognitive semantics: the many meanings of to run. In: Gries, Stefan Th. and Anatol Stefanowitsch (eds.). *Corpora in cognitive linguistics: corpus-based approaches to syntax and lexis*. Berlin, New York: Mouton de Gruyter.

von Eye, A. (1990). *Introduction to configural frequency analysis: The search for types and antitypes in cross-classification*. Cambridge: Cambridge University Press.

R Statistical Script

Gries, Stefan Th. (2009). *HCFA 3.2 – A Program for hierarchical configural frequency analysis for R for windows*.

Gries, Stefan Th. (2009). *BehavioralProfiles 1.01. A program for R 2.7.1 and higher*.

Travel assistance for this presentation is provided by
the J Gordin Kaplan Graduate Student Award at the University of Alberta

