

IBRAAM G. M. ABDELSAYED

## USO SCRITTO E PARLATO DELL'ARABO IN EGITTO: CARATTERISTICHE E DISTRIBUZIONE DEL LESSICO

### 1. PREMESSA

Molti studiosi si sono dedicati allo studio delle differenze caratterizzanti il parlato rispetto allo scritto, individuando delle caratteristiche specifiche di ciascuna forma (cfr. Biber 1988: 47). Si è messo così in evidenza come, in tutte le lingue, il parlato presenta delle caratteristiche tipiche che lo differenziano dalle altre forme e sistemi di significazione. In questo contributo ci focalizziamo, dunque, sulle specificità che caratterizzano il lessico del parlato rispetto a quello dello scritto, confrontando i dati emersi dalla ricerca sul primo *Lessico di frequenza dell'Arabo Parlato in Egitto*<sup>1</sup> con quelli attestati in Buckwalter e Parkinson (2011)<sup>2</sup>. Quest'ultimo è basato principalmente su testi scritti. Nella sezione successiva cercheremo di mettere in evidenza alcune differenze metodologiche essenziali tra il LAPE e il FDA prima di proseguire con il confronto. Va sottolineato che l'analisi discussa di seguito è di prima presen-

---

1 D'ora in avanti si usa l'acronimo LAPE. Si tratta della mia ricerca di dottorato svolta (e discussa il 6 febbraio 2019) presso l'Università per Stranieri di Siena, sotto la supervisione del professor Massimo Vedovelli.

2 D'ora in avanti FDA.

tazione approssimativo vista la mole di informazioni ricavabili dal LAPE e vista la disponibilità del FDA soltanto in versione cartacea.

## 2. LAPE E FDA A CONFRONTO: DIFFERENZE METODOLOGICHE E LIMITI

Anzitutto, va messo in luce che il FDA non si tratta tanto di un lessico di frequenza propriamente detto, quanto di un dizionario che mira a raccogliere e descrivere le parole arabe, fornendo qualche dato statistico riguardante il numero di occorrenze dei lemmi, ma non delle parole-forma. Infatti, i dati statistici sistematici del FDA non sono stati pubblicati, sebbene spendendo un po' di tempo, pochi dati possono essere ricostruiti. Mentre il LAPE è costituito da testi orali, quasi interamente spontanei o semi-spontanei (cfr. Voghera 1992), e perlopiù da interazioni comunicativi, il FDA è costituito invece per la maggior parte da testi scritti (circa il 90% e soltanto il 10% da testi parlati).

Il confronto che intendiamo operare non è da considerarsi esaustivo per diverse ragioni. La prima riguarda l'ampia differenza tra la dimensione di entrambi i corpora: circa 30 milioni di occorrenze di forme di lemmi per il FDA e 70000 per il LAPE. Mentre il LAPE è caratterizzato per essere una ricerca di natura prevalentemente qualitativa, il pregio principale del FDA consiste nella sua ampiezza numerica. Questa prima ragione ha implicato una differenza sostanziale a proposito della metodologia di cui ci si è avvalso in ciascuno di questi lavori e dei criteri di cui ci si è attenuto nello svolgimento delle cui fasi, e soprattutto nella lemmatizzazione e categorizzazione grammaticale. Bisogna, infatti, sottolineare che mentre per il LAPE le fasi di codifica sono state eseguite quasi completamente in maniera manuale sull'intero corpus, nel FDA, invece, sono state eseguite prevalentemente in maniera automatica. Considerata l'ampia dimensione del FDA e l'impossibilità di controllare in maniera completamente manuale un numero così vastissimo, gli autori del FDA hanno deciso di proseguire in maniera diversa. Buckwalter e Parkinson hanno iniziato con il calcolo di frequenza delle parole-forma grezze, così come erano scritte senza nessuna disambiguazione né segmentazione, cioè «to generate a frequency count of all surface forms (i.e. Arabic words as they are actually written, with prefixes and suffixes) taking into account not only the raw frequency of these surface forms but also their distribution over the corpus» (Buckwalter - Parkinson 2011: 4). Dopodiché si sono proceduti a determinare le prime 20000 parole-forma più frequenti per i testi scritti e le prime 10000 per i testi parlati. Soltanto queste 30000 parole-forma sono state lemmatizzate e taggate avvalendosi di Buckwalter morphological analyzer (2004). Gli autori sottolineano che ciò è stato eseguito in «two separate processes, one for the written corpus and one for the spoken corpus» (ibid.: 4). Da queste 30000 parole-forma sono stati ricavati 9900 lemmi. Già da questo dato emerge una prima ed evidente divergenza tra il FDA e il LAPE per quanto concerne la densità lessicale

in entrambi. Mentre la densità lessicale indicata dal valore dell'indice Uber<sup>3</sup> nel LAPE è del 20,28% (dato molto comparabile a quello del LIP<sup>4</sup>, pari al 21,61%), nel FDA lo stesso indica un grado molto elevato di diversità lessicale, arriva al 41,63%, quasi il doppio di quello del LAPE. Tenteremo nelle sezioni successive qualche ipotesi per interpretare questa rilevante divergenza.

Malgrado i criteri e le norme di lemmatizzazione e categorizzazione grammaticali del FDA non sono stati espressamente dichiarati, ciò nonostante sono facilmente ricavabili per analogia. Cercheremo nei seguenti punti di mettere in evidenza con degli esempi (tra i numerosi presenti) qualche limite manifestato nel FDA.

(1) Sono portate a due lemmi distinti parole-forma dello stesso lessema che sarebbero da portare allo stesso lemma, come ad esempio:

- il lemma عَلَى /'alāy/ 'su' (ibid., p. 10) e il lemma أ /'a/ 'su' (ibid.: 33);
- il lemma بِ /bi/ 'con' (ibid., p. 10) e il lemma بِي /bī/ 'con' (ibid.: 130).

(2) Non sono distinte funzioni diverse di un solo lessema ma sono portate allo stesso lemma al posto di portarle a due lemmi distinti, come ad esempio:

- il lemma رَئِيس /ratīs/ sia come sostantivo 'presidente' sia come aggettivo 'principale' (ibid.: 13);
- il lemma ضِدًّا /ḍidda/ sia come sostantivo 'contrario' sia come preposizione 'contro' (ibid.: 40).

(3) In diversi casi sembra mancare un'omogeneità nel criterio di lemmatizzazione, in particolar modo nel caso delle forme dialettali dello stesso lemma.

- Alcune volte, la forma dialettale e la forma "tradizionalmente" standard dello stesso lessema vengono portate a unico lemma, come ad esempio: la forma standard نَحْنُ /naḥnu/ 'noi' e le forme dialettali dello stesso lemma إِحْنَا /iḥnā/ (tipica dell'uso parlato in diversi paesi arabofoni, come in Egitto) e نِيْحْنَا /niḥnā/ (tipica dell'area del Levante). Nel FDA, tutte e tre le forme sono portate allo stesso lemma sotto la forma tradizionalmente standard (ibid.: 17).

- Altre volte, e in casi analoghi, si procede nella direzione contraria senza una chiara ragione, cioè sia la forma dialettale sia quella standard vengono portate invece a due lemmi separati, come ad esempio: la distinzione tra la forma tradizionalmente standard أَنَا /'anā/ 'io' (ibid.: 10) e la forma dialettale dello stesso lessema أَنِي /'ānī/ 'io'

---

3 È l'indice proposto da Dugast (1979) per calcolare la densità lessicale in riferimento al rapporto tra il numero di lemmi (o *types*) e quello di occorrenze (o *tokens*). Esso si ottiene dividendo il quadrato del valore logaritmico della N (il numero di occorrenze) per la differenza tra il valore logaritmico della V (il numero di lemmi) e quello della N, come mostra la formula:  $Uber\ index = (\log N)^2 / (\log N - \log V)$ .

4 L'acronimo sta per *Lessico di frequenza dell'Italiano Parlato* (De Mauro et. al. 1993).

(tipica dell'area dell'Iraq e Algeria) (ibid.: 187).

(4) In alcuni casi si evince un'inadeguatezza del lemma assegnato. È il caso di lemmi appartenenti a categorie grammaticali inconsuete per la tradizione linguistica araba (in particolare la categoria degli avverbi e delle interiezioni) e di conseguenza vengono aggregati insieme ad un altro lemma (spesso al sostantivo dello stesso lessema), trascurando la marca morfologica che determina tale uso, come ad esempio:

- il lessema أَهْلًا /ahl-ān/ 'benvenuto' o 'ciao' usato come interiezione viene aggregato al sostantivo أَهْل /ahl/ 'famiglia' o 'parente' (ibid.: 20). Tuttavia, la funzione interiezionale non può prescindere dalla marca morfologica /-ān/. Perciò emergerebbe, a nostro giudizio, necessario portare tale lessema a un lemma distinto dal lemma sostantivale أَهْل /ahl/ e assegnargli l'etichetta 'interiezione'.

(5) In altri casi invece si arguisce un'inadeguatezza della categoria grammaticale assegnata. È il caso di lemmi che, indipendentemente dalla funzione che assumano, vengono tradizionalmente analizzati come sostantivi, come ad esempio:

- il lessema بَعْض /ba'd/ 'poco' o 'alcuno' viene tradizionalmente classificato come sostantivo, a prescindere dalle funzioni<sup>5</sup> che possa assumere. Nel FDA a tale lemma viene assegnata l'etichetta di un sostantivo, sebbene l'esempio portato dagli autori stessi mostra che il lemma è usato come aggettivo indefinito (ibid.: 13).

### 3. LE INCONGRUENZE TRA LAPE E FDA: CARATTERISTICHE DEL PARLATO

Come abbiamo accennato in precedenza, nel FDA non sono state pubblicate tabelle statistiche di presentazione dei risultati. Non sono dunque disponibili i numeri statistici circa le categorie grammaticali né le classi di frequenza né le fasce d'autocopertura. Al fine di rendere eseguibile tale analisi abbiamo proceduto inizialmente alla digitalizzazione dei primi cento lemmi del FDA, creando una lista contenente rango d'uso, lemma, frequenza complessiva e percentuale di distribuzione. Una volta è stata realizzata questa lista, è diventato possibile confrontare il LAPE con il FDA. Riteniamo che il confronto basato soltanto sulla classe di frequenza (contenente i primi 100 lemmi) possa fornire dati molto rilevanti, poiché la linguistica statistica (Guiraud 1954: 10) rivela che tale classe copre normalmente circa il 60% di un dato corpus.

5 I dati del LAPE mettono in evidenza che possa assumere funzioni diverse: aggettivo indefinito (ad es.: "هؤلكو بعض الاسرار" /ha'ul:ukū ba'd 'il-'asrār/), pronome (ad es.: "انا وخالد نعرف بعض" /'anā wi-ḥālid ni'raf ba'd/) o sostantivo (ad es.: "البعض منكم شارك" /'il-ba'd minkum šārik/). Ci teniamo a sottolineare che tali parti del discorso, per noi, sono intese come categorie descrittive (o metalinguistiche) e non possono essere considerate dunque come realistiche. Assegnare a un'unità linguistica funzionale una specifica 'parte del discorso', significa cercare di individuare la funzione o il valore che essa assume nell'uso concreto e sceglierne tra le categorie disponibili quella che descriva meglio le cui caratteristiche.

Un primo confronto tra il LAPE e il FDA riguarda le categorie grammaticali rappresentate nei primi 100 lemmi. Alla classe di frequenza che comprende i primi 100 lemmi la distribuzione delle categorie grammaticali da un'ottica macroscopica mostra apparentemente una relativa uniformità: in entrambi i corpora 52 tra i primi 100 sono costituiti da parole grammaticali (dette anche 'vuote', 'sinsemantiche' o 'atematiche'; in inglese *synsemantic words*; in francese si parla di parole di *classe fermée*). Tuttavia, un'analisi più attenta può rivelare un alto grado di divergenza. Una prima divergenza riguarda la comparsa nel LAPE ma non nel FDA di nuove categorie grammaticali in posizioni molto avanzate, come la categoria di interiezione (5 lemmi) e di fonosimbolo (2 lemma). Una seconda divergenza concerne la categoria dei sostantivi. Mentre sui primi 100 lemmi nel LAPE si possono individuare soltanto 9 sostantivi, nel FDA se ne possono individuare 33: un dato molto significativo che comprova, a nostro parere, il forte assoggettamento dell'analisi fornita nel FDA dalla tradizione lessicografica araba che ritiene che le parti del discorso sono principalmente tre (nome, verbo e particella). Una altra incongruenza tra il LAPE e il FDA non concerne tanto le categorie grammaticali quanto i vocaboli stessi che ricorrono in questa classe di frequenza. Su 100 vocaboli soltanto 41 (cfr. tab. 1) in entrambi le liste corrispondono, benché occupino posizioni svariate nelle due liste.

Crediamo che le incongruenze appena esaminate tra il LAPE e il FDA possano essere interpretate – almeno in gran parte – in termini di caratteristiche specifiche della modalità parlata, quali cerchiamo di esporre nei paragrafi successivi.

### *3.1 Basso grado di densità lessicale*

Il lessico della lingua parlata in Egitto è tendenzialmente povero e di bassa densità lessicale, vale a dire è composto perlopiù da un numero relativamente ristretto di parole che coprono una percentuale di frequenza alta. Alla classe di frequenza che comprende i primi 100 lemmi, la percentuale di copertura arriva al 62,48% nel LAPE, mentre nel FDA è del 53,93%. Il dato indica un aumento rilevante della differenza d'autocopertura dei primi 100 lemmi, arriva al 8,55% a favore del LAPE. Eppure, le percentuali emerse dal LAPE risultano abbastanza plausibili con ciò che è stato sostenuto dai modelli teorici. Guiraud rileva che i primi 100 coprono normalmente circa il 60% di un dato corpus (Guiraud 1954: 10). Anche se prendessimo in considerazione il fatto che la percentuale nel LAPE sfiori del 2,48% quella individuata da Guiraud, quella emersa nel FDA si abbassi del 6,07%: dato, comunque, non irrilevante.

A nostro parere, tale divergenza oltre a rispecchiare una caratteristica specifica del parlato, si deve anche al fatto che il corpus FDA include varietà linguistiche che sono nettamente distinti: testi scritti (costituiscono il 90% del corpus) e parlati (il 10%) provenienti da 20 aree geo-linguisticamente diverse (Algeria, Bahrain, Egitto, Golfo, ecc.) (Buckwalter - Parkinson 2011: xi). Ciò ha portato inevitabilmente a una maggiore ripartizione in lemmi distinti di unità lessicali sinonimiche e funzionalmente interscambiabili. Tuttavia, bisogna precisare che tale interscambiabilità è fortemente

USO SCRITTO E PARLATO DELL'ARABO IN EGITTO

delimitata non soltanto dalle consuete variabilità diamesiche e diatopiche, ma anche e soprattutto da una variazione del codice linguistico stesso.

VOCABOLO LEMMA	C. G.	GLOSSA	RANGO D'USO	
			LAPE	FDA
ال /'al/	part_det	'il'	1	1
و /wa/	conj	'e'	2	2
هو /huwā/	pron	'lui'	3	20
ل /li/	prep	'a' o 'per'	4	5
في /fī/	prep	'in'	5	3
انا /'anā/	pron	'io'	6	25
ب /b-/	prog	'sta + gerundio'	7	31
هي /hiyā/	pron	'lei'	8	33
أنت /'anta/	pron	'tu' [maschile]	10	54
ب /bi/	prep	'con' [mezzo]	11	6
نحن /naḥnu/	pron	'noi'	12	97
أن /'an/	conj	'che'	13	13
على /'alaÿ/	prep	'su'	14	7
كان /kāna/	verb	'era'	16	10
من /min/	prep	'di'	18	4
قال /qāla/	verb	'dire'	19	15
كل /kull/	adj / noun	'tutto', 'ogni' o 'intero'	23	19
ف /fa/	conj	'poi'	25	21
يا /yā/	part_voc	'o' [Signore]	27	35
مع /ma'a/	prep	'con' [persona]	28	17
الله /'allāh/	noun_prop	'Allah'	32	12
او /'aw/	conj	'oppure'	36	23
عند /'ind/	prep	'da'	41	40
أي /'ayy/	adj / noun	'qualsiasi' o 'ogni'	44	46
لو /law/	conj	'se'	46	87
أحد /'aḥad/	noun / pron	'qualcuno'	54	63
عرف /'arif/	verb	'conoscere'	58	49
واحد /wāḥid/	num / pron	'uno'	60	56
لان /li'āna/	conj	'perché'	61	57
يوم /yawm/	noun	'giorno'	63	26
بعد /ba'd/	prep	'dopo'	67	34

عن	/ʿan/	prep	‘riguardo’	72	14
نفس	/nafs/	adj / noun	‘stesso’	73	44
كثير	/kaṭīr/	adj	‘molto’	74	55
لكن	/lākin/	conj	‘ma’	75	91
اول	/ʾawal/	num / adj	‘primo’	84	41
سنة	/sana[t]/	noun	‘anno’	85	69
رجل	/raǧul/	noun	‘uomo’	90	92
	/mawḏūʿ/	noun	‘argomento’	95	95
لا	/lā/	part_neg	‘non’	98	11
بلد	/balad/	noun	‘paese’	99	99

**tab. 1.** *Vocaboli corrispondenti nel LAPE e nel FDA.*

Nel caso della lingua araba, lo spostamento areale produce variazione lessicale non solo nell’uso parlato e quello scritto della stessa lingua (ciò è un fenomeno comune in tutte le lingue), ma implica anche l’entrata in gioco di codici linguisticamente distinti. Per spiegare meglio ciò che intendiamo dire, riportiamo nella tab. 2, soltanto per fare un esempio, i lessemi attestati nel FDA per dire ‘che’ o ‘quale cosa’.

LESSEMA	RANGO D’USO NEL FDA	AREA GEOGRAFICA DI RILEVAZIONE <sup>6</sup>
شو /šū/ (ibid.: 21)	141	Levante, Kuwait, Emirati Arabi Uniti
ما /mā/ (ibid.: 29)	231	-
ماذا /māḏā/ (ibid.: 33)	275	-
إيه /ʾīh/ (ibid.: 45)	405	Egitto
إيش /ʾīš/ (ibid.: 104)	1.053	Levante e Golfo
شو & ش /šū/ (ibid.: 132)	1.364	Iraq
شونو /šinū/ (ibid.: 134)	1.390	Golfo e Iraq
واش /wāš/ (ibid.: 421)	4.835	Maghreb (paesi del Nordafrica)

**tab. 2.** *Distribuzione di lemmi sinonimici ma appartenenti a codici linguistici diversi nel FDA.*

Come si può desumere dalla tabella, lo spostamento da un’area geografica ad un’altra ha portato alla ripartizione in otto lemmi distinti di unità lessicali sinonimiche e funzionalmente interscambiabili. Vista la percentuale comunque ridotta dei testi parlati inclusi nel FDA (solo il 10%), la presenza di parole marcate diatopicamente (se le vogliamo chiamare così) tra le prime 1.000 conferma il fatto che esse non dipendano, come si pensa spesso, dall’uso parlato della lingua, ma indubbiamente esse

<sup>6</sup> La mancanza di una specificazione areale indicata con il trattino ‘-’ significa che il lessema è tipico dell’arabo standard.

sono pienamente integrate anche nell'uso scritto; altrimenti non sarebbero arrivate a posizioni così avanzate nella lista. Basti pensare al primo lessema nella (cfr. tab. 2) *شو* /šū/ che si ritiene che sia marcata diatopicamente ma allo stesso tempo più frequente della sua variante "standard" *ما* /mā/. Infatti, ciò è stato affermato da Buckwalter - Parkinson (2011: 5) che scrivono:

Because the spoken corpus accounted for only 10 percent of the data, much of the high-frequency vocabulary of spoken colloquial Arabic did not make it into the top 5,000 list. Those words that did make it, however, were dialect vocabulary items that are now also widely used in written media, and were observed to have relatively high frequencies and good distribution especially in the "forums" section of the corpus.

Invece nel LAPE al contrario di queste otto parole ne compiano solo tre: *إيه* /'ih/ (rango d'uso *ما* 29 /mā/ rango d'uso 481) e *ماذا* /mādā/ (rango d'uso 2.467). Questi tre lessemi, oltre ad essere funzionalmente interscambiabili, sono in una relazione di distribuzione libera sull'asse paradigmatico, o possiamo dire, adattando i termini della fonetica articolatoria al nostro discorso, sono varianti libere a disposizione del parlante che ne seleziona liberamente tra di loro. Tuttavia, tale libertà distribuzionale ha una funzione sociolinguistica ben precisa, ovvero quella di determinare il grado di formalità dell'enunciato, adeguandolo al contesto in cui si svolge, al tipo di relazione stabilita con i partecipanti all'interazione comunicativa e al canale tramite cui passa l'enunciato.

Al contrario, la relazione che si stabilisce tra queste tre lessemi e gli altri cinque mostrati è completamente di ordine diverso; non è più una relazione tra varianti libere sull'asse paradigmatico dello stesso codice linguistico, bensì di codici diversi appartenenti a comunità di parlanti distinti. L'entrata in gioco di questi altri cinque lessemi fa venir meno questo meccanismo di libertà distribuzionale, perché la variazione, in questo caso, supera i confini della variabilità di un solo codice, facendo entrare in gioco più codici. Anche se risulterebbero funzionalmente interscambiabili, tale interscambiabilità non potrà funzionare senza una modificazione del codice stesso e della società in cui verrà usato, e dunque degli attori stessi dello scambio comunicativo.

### *3.2 Alto uso di unità lessicali polisemiche*

Il lessico del LAPE è caratterizzato da un alto uso di lessemi polisemici (cioè semanticamente ampi), piuttosto che lessemi con riferimenti semantici più specifici. Ciò si deve al fatto che, quando si parla, l'elaborazione dell'enunciato coincide con quella della progettazione. Data la mancanza di tempo sufficiente per svolgere al contempo queste due operazioni, si fa ricorso a lessemi, caratterizzati da una struttura semanticamente ampia, meno specifica e meno elaborata. Basti pensare al sostantivo *حاجة* /hāga[t]/ 'cosa' che emerge dai dati del LAPE come il sostantivo più frequente nel parlato egiziano al 24° posto, mentre occupa il 669° posto nella lista del FDA. Un altro



caso è il sostantivo ناس /nās/ ‘gente’ che occupa il 37° posto nel LAPE, mentre viene incorporato al sostantivo إنسان /’nsān/ ‘essere umano’ al 204° posto del FDA. Analogamente anche i verbi: عَمِلَ /’amila/ ‘fare’ o ‘lavorare’ (al 31° posto nel LAPE e al 137° posto del FDA), حَلَّى /halla’/ ‘lasciare’, ‘permettere’, ‘far diventare’ ecc. (al 78° posto nel LAPE e al 763° nel FDA) e جَاب /gāb/ ‘espandere’, ‘portare’, ‘condurre con sé’ ecc. (al 88° posto nel LAPE e al 4.991° nel FDA) e طَلَعَ /ṭala’a/ ‘diventare’, ‘comparire’, ‘emergere’, ‘uscire’ ecc. (al 94° posto nel LAPE e al 306° nel FDA). Tra i primi 100 lemmi del FDA compaiono invece lessemi semanticamente più specifici, come ad esempio: دولة /dawla[t]/ ‘stato’ o ‘governo’ (al 51° posto nel FDA e al 159° nel LAPE), قُوَّة /quw:a[t]/ ‘potenza’ (al 71° posto nel FDA e al 508° nel LAPE) e وَزِير /wazīr/ ‘ministro’ (al 96° posto nel FDA e al 245° nel LAPE).

### 3.3 Alto uso di unità lessicali polifunzionali

Sembra che nell’uso parlato dell’arabo in Egitto viene data una certa preferenza per unità lessicali polifunzionali. È il caso dei pronomi relativi che ricoprono 4 posti tra i primi 100 nel FDA: الَّتِي /’alladī/ [s./f.] al °18 posto, الَّذِي /’alladī/ [s./m.] al °24 ما /mā/ [invariabile] al 30°, مَنْ /man/ [invariabile] al 30° posto; mentre nel LAPE per la funzione ‘pron\_rel’ viene ad occuparsi soltanto una posizione tra i primi 100 dal lessema polifunzionale الِلي /’illi/ [invariabile per i tratti di genere e numero] al 15° posto. Lo stesso vale per le particelle di negazione (لَا /lā/, لَمْ /lam/ e مَا /mā/) che – sebbene si differenzino funzionalmente nella lingua scritta – occupano nel FDA rispettivamente i posti 11°, 27° e 28°, mentre i dati del LAPE mostrano come nel parlato viene data preferenza alla particella di negazione polifunzionali – poiché possa sostituire le tre sopracitate – مِشْ /miš/ che ricopre il 9° posto.

### 3.4 Uso di lessico con maggior grado di neutralità e minore grado di marcatezza

Tramite l’analisi dei primi 100 lemmi del corpus LAPE e quelli del FDA si è rilevato come nell’uso parlato in Egitto non siano presenti idiomi contrapposti, ma si avverta un continuum linguistico unitario (sebbene eteroglossico) che costituisce il sistema arabo-egiziano inteso come “*Sprachesystem*” (cfr. Coşeriu 1977: 115). Con ciò non si vuole trascurare la presenza non tanto di altre varietà linguistiche (considerate alte o standard), quanto piuttosto delle loro variabili dovute alle dimensioni di variazione (diamesica, diafasica, diastratica e diatopica). Tale ipotesi non individua alcun polo di questo continuum, bensì un centro di neutralità, all’interno del quale prevale una tecnica del discorso sintopica (cioè senza variazione nello spazio), sinstratica (cioè senza variazione relative alle stratificazioni sociolinguistiche e socioculturali) e sinfasica (cioè senza variazione relative al tema del testo né al rapporto con i partecipanti all’interazione comunicativa), cioè un insieme soggiacente di norme non marcate, testimoniate dalle loro occorrenze osservabili (poiché ricorrono nei testi con frequenze relativamente stabili). Ad esempio, i lemmi del LAPE عَشَانَ /’ašān/ ‘per’ (al

°49 posto), شَاف /šāf/ ‘vedere’ (al 52° posto) e إِزَاي /'izzay/ ‘come’ (al °79 posto) sono confrontabili, rispettivamente, con i lemmi del FDA حَتَّى /ḥattay/ ‘per’ (al 81° posto), رَأَى /ra'aÿ/ ‘vedere’ (al 75° posto) e كَيْفَ /kaīfa/ ‘come’ (al 67° posto). Sebbene i lemmi del LAPE (عَشَان /'ašān/, شَاف /šāf/ e إِزَاي /'izzay/) siano considerati dagli studi sociolinguistici come dialettali, essi risultano – se considerata la loro frequenza e la loro relativa uniformità distribuzionale nei sottocorpora – meno marcate rispetto a quelli del FDA (حَتَّى /ḥattay/, رَأَى /ra'aÿ/ e كَيْفَ /kaīfa/). Si ipotizza che la scelta di questi ultimi sia fortemente condizionata dal canale di trasmissione del testo (più usati nello scritto), ma anche e soprattutto dal grado di formalità del testo (più frequenti nei testi formale). Ciò appare ancora più evidente se andassimo ad osservare la distribuzione della loro di frequenza nei sottocorpora del LAPE (cfr. tab. 3).

TOT	GR V	GR IV	GR III	GR II	GR I	C. G.	LEMMA
156	24	27	50	30	25	prep	'ašān عَشَان
2	0	0	1	1	0	prep	ḥat:y حَتَّى
142	33	20	23	34	32	verb	šāf شَاف
19	0	16	3	0	0	verb	rā'y رَأَى
112	27	17	11	27	30	interrog	'iz:āy إِزَاي
7	0	7	0	0	0	interrog	kayfa كَيْفَ

tab. 3. Distribuzione di frequenza e marcatezza di alcuni lemmi nel LAPE.

Come si può vedere nella tabella i lemmi عَشَان /'ašān/, شَاف /šāf/ e إِزَاي /'izzay/ sono distribuiti in maniere più omogenei all'interno dei sottocorpora – ciò rispecchia la loro maggiore neutralità – rispetto agli altri tre lemmi حَتَّى /ḥattay/, رَأَى /ra'aÿ/ e كَيْفَ /kaīfa/ che sono concentrati nei sottocorpora caratterizzati da un tipo di parlato meno spontaneo e più formale, e in particolare nel Gruppo III (testi appartenenti alla tipologia testuale “scambio comunicativo bidirezionale con presa di parola non libera, faccia a faccia”, come ad esempio “assemblee”, “dibattiti” ecc.) e nel Gruppo IV (testi appartenenti alla tipologia testuale “scambio comunicativo monodirezionale in presenza del/i destinatario/i”, come ad esempio, “comizi politici”, “omelie religiose” ecc.).

### 3.5 Maggior ricorso a elementi deittici

Dal confronto tra il LAPE e il FDA sembra che il parlato sia caratterizzato per un maggiore ricorso a elementi intrinsecamente *deittici*, come i pronomi personali. Mentre nel caso dello scritto i pronomi funzionano da *anafora*, poiché la loro interpretazione richiede un accesso soltanto al contesto linguistico, cioè al co-testo, nel caso del parlato i pronomi fungono piuttosto da *deittici* poiché la loro corretta interpretazione non è limitata al co-testo ma richiede un accesso di tipo extralinguistico, ovvero al contesto situazionale. Il confronto tra il LAPE e il FDA ha mostrato un forte

slittamento nella posizione che occupano i pronomi, la quale è molto più avanzata nel LAPE rispetto al FDA: هُوَ /huwā/ 'lui' (al 3° posto nel LAPE e al 20° nel FDA), أَنَا /'anā/ 'io' (°6 nel LAPE e °25 nel FDA), هِيَ /hiyā/ 'lei' (8° nel LAPE e 33° nel FDA), أَنْتَ /'anta/ 'tu' [maschile] (°10 nel LAPE e °54 nel FDA), نَحْنُ /nahnu/ 'noi' (12° nel LAPE e 97° nel FDA). I testi scritti non sono ancorati né al luogo né al tempo dell'evento comunicativo: tendono perciò a essere decontestualizzati, e dunque i riferimenti anaforici devono essere esplicitati e resi con riferimenti puntuali. A differenza dei testi orali, che sono lineari e non-correggibili, quelli scritti possono essere letti e riletti dapprima dallo stesso autore, che può intervenire con correzioni e rifacimenti, e poi dal lettore, che può leggerlo per intero o diviso in parti e rileggerne i passi interessanti o complicati. Infatti, i deittici – come i pronomi personali nel parlato – coinvolgono due realtà di ordine diverso: una di ordine linguistico, dunque interna all'enunciato, e un'altra di ordine extralinguistico, la quale va cercata nel contesto situazionale in cui viene svolto l'enunciato. I deittici dipendono primariamente da ciò che Bühler (1965: 102) chiama il *centro deittico* (in tedesco '*deiktische Zentrum*') o *origo* che coincide di solito con il parlante che rappresenta il punto di riferimento verso cui sono orientati i deittici. Oltre ai deittici personali, i dati del LAPE assegnano una posizione molto avanzata ai deittici temporali rispetto alla posizione che occupano nella lista del FDA, come ad esempio: إِنِّهَارًا /'in-nahardā/ 'oggi' (al °80 posto nel LAPE e al 4.969° nel FDA) e دِلْوَأْتِي /dilwa'ti/ 'adesso' (al 87° posto nel LAPE e al 4.590° nel FDA). Tale dato conferma la centrale specificità dei deittici nella modalità interattiva del parlato.

### 3.6 Maggior ricorso a segnali discorsivi

Il parlato è caratterizzato da un maggiore utilizzo dei segnali discorsivi (in inglese *discourse markers*) (cfr. Schiffrin 1987). A tale riguardo afferma Voghera che «quando si parla si devono non solo fornire informazioni sufficienti perché il contenuto che vogliamo trasmettere arrivi al destinatario, ma anche fornire una griglia per l'elaborazione del testo» (2001: 81). Tra i segnali discorsivi più frequenti nel LAPE rispetto al FDA possiamo ricordare le interiezioni: يَعْني /ya'nī/ 'cioè' (al 17° posto nel LAPE e al 751° nel FDA<sup>7</sup>), لا /lā/ 'no' (al 35° posto nel LAPE e assente nel FDA), بَعِي /ba'ÿ/ 'quindi' o 'dunque' (al 42° posto nel LAPE e al 3.458° nel FDA<sup>8</sup>), طَيِّب /ṭayyb/ 'va bene' (al 42° posto nel LAPE e al 147° nel FDA<sup>9</sup>) e مَا /mā/ 'proprio' (al 83° posto nel

7 Nel FDA è stata assegnata l'etichetta *part* 'particella'.

8 Anche in questo caso è stata assegnata l'etichetta *part* 'particella'.

9 La posizione apparentemente alta si deve al fatto che nel FDA non si sia distinto la funzione aggettivale del lemma طَيِّب /ṭayyb/ 'buono' (che ha come caratteristica l'ammissibilità di introdurre l'articolo determinativo e della declinazione per genere e numero) da quella interiezionale (che ha come caratteristica specifica l'essere una parte invariabile del discorso). Invece, nel LAPE si porta a lemma diverso ogni funzione (l'aggettivo طَيِّب /ṭayyb/ agg. 'buono' ricorre al 874° posto).

LAPE e al 1.837° nel FDA<sup>10</sup>); i fonosimboli: /ah/ 'sì' o 'esatto' (al 38° posto nel LAPE e al 142° nel FDA) ed /eh/ 'riempitivo' – serve a mantenere il turno – (al 56° posto nel LAPE e al 164° nel FDA); e infine gli avverbi modali: كِدَا /kidā/ 'così' (al 33° posto nel LAPE e al 1.000° nel FDA), طَبِعًا /ṭab'-'ān/ 'naturalmente' o 'ovviamente' (al 64° posto nel LAPE e al 340° nel FDA<sup>11</sup>) e بَرِّضُو /barḍū/ 'anche', 'pure' o 'persino' (al 82° posto nel LAPE e al 2.715° nel FDA).

Owens e Rockwood (2008: 85) sottolineano che «[t]he detailed study of discourse markers in spoken Arabic is largely in its infancy». Da questo punto di vista, il corpus LAPE costituisce un prezioso strumento per poter condurre future ricerche su elementi specifici del parlato egiziano, come appunto i segnali discorsivi, poiché offre dati empirici ed esempi concreti della lingua parlata effettivamente nel contesto egiziano.

#### 4. CONCLUSIONI E PROSPETTIVE

Dal confronto tra il LAPE e il FDA è emerso che – sebbene vi siano importanti incongruenze interpretabili in termini di caratteristiche specifiche della modalità parlata – la lingua scritta (la cui fonte primaria è l'arabo standard) e la lingua parlata (il cosiddetto dialetto egiziano) in Egitto condividono aspetti non irrilevanti. Tuttavia, non bisogna correre il rischio di perdere di vista che il confronto descritto nel presente contributo fosse tra lemmi che di per sé sono unità astratte. Perciò bisogna ricordare che il processo di lemmatizzazione, alla base di entrambi i lavori (LAPE e FDA), consista nel portare la forma concreta – spesso marcata da diverse dimensioni di variazione (diafasica, diastratica, diatopica e/o diamesica) – a una forma astratta non marcata (o cosiddetta standard). Le norme standard della lingua araba sono, ormai da secoli, descritte – se non addirittura prescritte; di conseguenza, anche noi ci siamo trovati, in questa ricerca, a dover standardizzare il lemma in tutti i casi in cui la parola-forma in esame, nonostante le modifiche (a livello fonetico-fonologico, morfologico ecc.) subite rispetto alla forma standard, derivasse da essa.

Al fine di individuare le tendenze generali che caratterizzano la lingua contemporanea in Egitto, pertanto, occorrerà procedere nella direzione opposta a quella della lemmatizzazione, ossia partire dalle parole-forma, esaminando in che misura si differenzino dal lemma standardizzato, e in che misura restino tuttora ancorate alle

10 Nel FDA è analizzata come particella di ridondanza.

11 Anche in questo caso la posizione apparentemente alta si deve a una mancanza di precisione nell'analisi del FDA che mette insieme la funzione sostantivata del lemma طَبِعَ /ṭab' / 'carattere' o 'temperamento' (che ha come caratteristica l'ammissibilità di introdurre l'articolo determinativo e della declinazione per numero) assieme a quella avverbiale طَبِعًا /ṭab'an/ 'naturalmente' o 'certamente' (che ha come caratteristica specifica l'introduzione della marca morfologica della nunazione detta anche تَنْوِين /tanwīn/ e l'essere una parte invariabile del discorso). Nel LAPE il sostantivo طَبِعَ /ṭab' / 'carattere' o 'temperamento' è assente.

caratteristiche del modello storicamente standardizzato.

## BIBLIOGRAFIA

- Bibber 1988 = Douglas Biber, *Variation across Speech and Writing*, Cambridge, Cambridge University Press.
- Buckwalter - Parkinson 2011 = Tim Buckwalter - Dilworth Parkinson, *A Frequency Dictionary of Arabic*, London, Routledge.
- Bühler 1965 = Karl Bühler, *Sprachtheorie: Die Darstellungsfunktion der Sprache*, 2<sup>a</sup> ed., Stuttgart, Gustav Fischer (1<sup>a</sup> ed. 1934).
- Coşeriu 1977 = Eugenio Coşeriu, 'Historische Sprache' und 'Dialekt', in Joachim Göschel - Pavle Ivić - Kurt Kehr (a cura di), *Dialekt und Dialektologie. Ergebnisse des internationalen Symposiums, Zur Theorie des Dialekts*, Wiesbaden, Marburg & Lahn, pp. 106-122.
- De Mauro et. al. 1993 = Tullio De Mauro et alii, *Lessico di frequenza dell'italiano parlato*, Milano, ETAS.
- Dugast 1979 = Daniel Dugast, *Vocabulaire et Stylistique: I Théâtre et Dialogue, Travaux de Linguistique Quantitative*, Geneva, Slatkine-Champion.
- Guiraud 1954 = Pierre Guiraud, *Les caractères statistiques du vocabulaire: essai de méthodologie*, Paris, Presses universitaires de France.
- Owens - Rockwood 2008 = Jonathan Owens - Trent Rockwood, *yaʿni: what it (really) means*, in Dilworth Parkinson (a cura di), *Perspectives on Arabic Linguistics XXI: Papers from the Twenty-first Annual Symposium on Arabic Linguistics, Provo, Utah, March 2007*, Amsterdam & Philadelphia, John Benjamins B. V., pp. 83-113.
- Schiffrin 1987 = Deborah Schiffrin, *Discourse Markers*, Cambridge, Cambridge University Press.
- Voghera 1992 = Miriam Voghera, *Sintassi e intonazione nell'italiano parlato*, Bologna, il Mulino.
- Voghera 2001 = Miriam Voghera, *Teorie linguistiche e dati di parlato*, in Federico Albano Leoni, - Stenta Krosbakken - Rosanna Sornicola - Carolina Stromboli (a cura di), *Dati empirici e teorie linguistiche. Atti del XXXIII Congresso Internazionale di Studi della Società di linguistica (SLI)*, Roma, Bulzoni, pp. 75-95.