

Fenomeno della collocabilità ristretta nell'italiano di oggi



Zora Obstová (Praga)

ABSTRACT

In tutte le lingue naturali troviamo parole sprovviste di autonomia sintattica e semantica che possono esistere soltanto all'interno di una combinazione lessicale. Queste parole, designate con i termini *cranberry words*, *bound words*, *unique words* o *monocollocable words* (*parole monocollocabili*, PM), sono caratterizzate da un raggio collocazionale estremamente ristretto (che va solitamente da 1 fino a ± 7 collocati). Il presente articolo vuole descrivere il fenomeno della monocollocabilità nell'italiano di oggi, basandosi sulle liste delle PM italiane estratte dai corpora e contenute nel libro *Language Periphery, Monocollocable Words in English, German, Italian and Czech* (Čermák et al., 2016). Le PM italiane e le locuzioni a cui esse fanno capo vengono analizzate sotto il profilo sintattico, semantico, collocazionale e frequenziale. La monocollocabilità è un fenomeno mutevole nel tempo: nonostante le PM vengano spesso considerate "reliquie del passato", i dati raccolti mostrano che una progressiva trasformazione di alcune parole con ampio raggio collocazionale in parole monocollocabili avviene anche nel lessico attuale.

PAROLE CHIAVE

collocazione unica, parole monocollocabili, locuzioni, polirematiche, corpora

ABSTRACT

There is a specific combinatorial periphery in any language consisting of words whose combinatorial potential is extremely restricted. These words, which are usually referred to as *bound words*, *unique words*, *cranberry words* or *monocollocable words* (MWs), belong to small and closed collocation paradigms, their number of collocates ranging from one to a few (usually ± 7). The present article tries to describe the phenomenon of monocollocability in Italian, basing the analysis on a list of Italian MWs extracted from corpora and contained in the book *Language Periphery, Monocollocable Words in English, German, Italian and Czech* (Čermák et al., 2016). Italian MWs and the fixed combinations in which they occur are analysed in terms of syntactic structures, semantic features, collocation structures and frequency. Monocollocability is a phenomenon subject to change in time: even though MWs are often considered to be "relicts of the past", the collected data prove that progressive restriction of the combinatorial capacity of certain words can be observed in Present-Day Italian as well.

KEYWORDS

monocollocable words, cranberry words, multiword expressions, Present-Day Italian, corpora

1. CRANBERRY WORDS O PAROLE MONOCOLLOCABILI: DEFINIZIONE

Nelle nostre conversazioni, nei testi letterari, giornalistici, tecnici o pubblicitari che leggiamo ci troviamo spesso di fronte a parole che sono sprovviste di autonomia sintattica e semantica e che possono esistere soltanto all'interno di una combinazione lessicale. Locuzioni di questo tipo sono abbastanza frequenti nel nostro esprimerci quotidiano: cantiamo a *squarciagola*, corriamo all'*impazzata*, beviamo a *garganella*, andiamo a *braccetto*, stiamo all'*erta*, facciamo qualcosa a *malincuore* e nella *stragrande*



maggioranza dei casi ci riusciamo a *malapena*. *Impazzata* o *squarciagola* come parole autonome non esistono, anche se sono attestate come lemmi nei dizionari moderni. Tuttavia, data l'assenza di un proprio significato, spesso non hanno una definizione, ma compaiono affiancate dalla dicitura "usato solo nella loc." e simili. Non di rado queste espressioni sono relitti linguistici che una volta hanno avuto un'esistenza autonoma e maggiori possibilità combinatorie che sono andate riducendosi col passare del tempo cosicché esse oggi ci appaiono immotivate, semanticamente opache.

Queste parole, che probabilmente esistono in tutte le lingue naturali, vengono designate con i termini *cranberry words* (in analogia con *cranberry morphemes*, cfr. Bloomfield 1933 e Aronoff 1976), *bound words* o *unique words*; in tedesco si parla di *unikale Komponente* (cfr. Fleischer 1989) o *Unikalia* (cfr. Dobrovolskij-Piirainen 1994), in italiano di *componenti (vocaboli) a collocazione unica* (cfr. Veland 2004, 2006). Per il ceco è stato introdotto il termine *monokolokabilní slova* (*parole monocollocabili*, PM), coniato da Čermák (1982, 2014). Čermák definisce la *monocollocabilità* come un'anomalia collocazionale di una parola con un raggio collocazionale estremamente ristretto, formato da 1 fino a ± 7 collocati¹ (per i casi di omonimia e polisemia vedi *infra*). Non si tratta, dunque, nonostante il termine *mono*-collocabile possa suggerirlo, solo di parole con un collocato unico (tali parole sono, tra l'altro, molto rare). Visto che la monocollocabilità è un fenomeno scalare, la categoria può essere allargata anche a parole con più collocati fino a comprendere persino quelle forme che hanno sì un raggio collocazionale più aperto, ma che dimostrano, nell'uso attuale, una forte tendenza a prediligere un certo collocato rispetto ad altri — cioè a formare con un'altra parola, autosemantica o grammaticale, una *locuzione*.

2. METODI PER IDENTIFICARE LE PM

Le parole monocollocabili sono state finora studiate prevalentemente nel contesto della fraseologia (Dobrovolskij parla di "phraseologisch gebundenen Formative") dato che molte PM fanno parte di locuzioni idiomatiche. Per identificare parole a collocazione estremamente ristretta la ricerca si era finora basata su fonti diverse, utilizzando dizionari, questionari, interviste, analisi dei testi, introspezione. Dobrovolskij (1988) e Dobrovolskij e Piirainen (1994), per esempio, hanno creato un importante corpus di Unikalia esistenti in tedesco, inglese e olandese; le parole e le locuzioni da loro identificate hanno costituito la base per la successiva ricerca svolta in questo campo da Trawiński, Sailer, Soehn, Lemnitzer e Richter (cfr. CoDII). Scarseggiano, invece, studi dedicati alla monocollocabilità in italiano: tra i pochi contributi citiamo gli articoli di Veland (2004, 2006) che ha stilato due liste di vocaboli non autonomi (composte di 116 e di 90 unità), basandosi sul *Dizionario Garzanti di italiano* (DGI) e sul *Dizionario della lingua italiana* di De Mauro (2000) e registrando tutte le parole accom-

1 La nozione di collocato non va qui intesa nella concezione semantica di base — *collocato/collocatore* (cfr. Hausman 2004), bensì nei termini della linguistica dei corpora, come ogni lemma o forma flessa che co-occorre in modo significativo con una PM (cfr. i termini *node* — *collocate* coniatati da Sinclair).



pagnate dalla dicitura “solo nella loc.” (o “nella loc.”, “soprattutto nella loc.” ecc.); di queste parole e delle rispettive locuzioni, l’autore ha analizzato in particolare le proprietà sintattiche e collocazionali. Di *Unikalia* in italiano, considerati “tracce della diacronia nella sincronia”, si è occupata anche Konecny (2014) che ha analizzato alcune locuzioni idiomatiche con le PM, percepite soprattutto come risultati di un processo diacronico unicalizzante.

Esiste, comunque, anche un altro metodo per identificare le PM: la loro estrazione dai corpora, resa possibile dall’analisi basata sull’indice di Herfindahl-Hirschman (HHI), utilizzato per quantificare la diversità del contesto (cfr. Cvrček 2013, 119)². Questo metodo è stato applicato (cfr. Čermák, F., Čermák, J., Obstová, Z., Vachková, M., 2016) a corpora di riferimento di simile estensione in quattro lingue (ceco, italiano, tedesco e inglese; per l’italiano è stato utilizzato il CORIS che contiene 130 milioni di parole) e ha dato, per molti versi, risultati abbastanza sorprendenti.

L’approccio sistematico, applicato ai dati di quattro lingue tipologicamente diverse, permette di cogliere il fenomeno della monoccollocabilità in tutta la sua complessità. Le liste delle PM estratte dai corpora differiscono per molti aspetti dalle collezioni di *cranberry words* sopra menzionate. Queste differenze rispecchiano alcune specificità su cui l’analisi si è basata:

a) Gli autori del libro sostengono che la “monoccollocabilità” può caratterizzare il lessema ma più spesso riguarda “parole” intese come forme flesse (*word-forms*); per questo motivo la ricerca è stata effettuata su tutte le forme presenti nei corpora. Le conseguenze di questa scelta sono più tangibili nelle lingue con flessione nominale più sviluppata come il ceco dove alcune PM esistono, sotto il profilo sincronico, soltanto come forme flesse e non come lemmi (*jít k duhu*). In italiano il fenomeno riguarda prevalentemente gli aggettivi e alcuni nomi. In molti casi queste parole presentano un raggio collocazionale estremamente ristretto, se utilizzate al plurale. Come esempi possiamo citare le forme *paraggi* e *antipodi*, le quali — nonostante vengano registrate nei dizionari sotto i lemmi *paraggio* e *antipode* — compaiono nell’italiano attuale esclusivamente al plurale, nelle locuzioni *nei paraggi (di qc.)* e *agli antipodi*. Dal punto di vista sincronico sarebbe sbagliato, come osserva Čermák (2014, 12), estrarre questi costituenti dalla locuzione e attribuire loro determinate caratteristiche (lemma, parte del discorso, significato autonomo). Altre PM sono invece difettive del plurale, come la parola *impazzata* (nei dizionari accompagnata dalla dicitura *solo sing.*), utilizzata esclusivamente al singolare, nella locuzione *all’impazzata*. A volte la collocabilità ristretta è tipica soltanto per una forma del lemma: è il caso della parola *sbarrati*, che in base ai dati estratti dal corpus va considerata monoccollocabile (*occhi sbarrati*), mentre le altre forme del paradigma non presentano particolari limitazioni di collocabilità.

2 L’indice HHI è un mezzo statistico, usato in economia principalmente per misurare il grado di concorrenza presente in un determinato mercato. È dato dalla somma dei quadrati delle quote di mercato di tutte le imprese presenti nel settore (cfr. Hirschman 1964). Questo indice può essere utilizzato anche in linguistica (cfr. Cvrček 2013) per quantificare la diversità del contesto di una parola, ossia per stabilire se una determinata posizione può essere occupata da tante parole diverse o se invece una o più parole hanno “il monopolio” su quella posizione.



b) La seconda specificità del metodo utilizzato per estrarre le PM dai corpora consiste nell'allargamento del raggio collocazionale: mentre la ricerca precedente si era concentrata prevalentemente su parole a collocabilità unica o quasi unica (parole che ricorrono in una sola costruzione), i dati estratti dai corpora in base all'indice HHI corrispondono a un'accezione più ampia della monocollocabilità, dove il numero dei collocati varia, come menzionato sopra, da 1 a ± 7 . Inoltre non vengono esclusi frequenti casi di omonimia o polisemia (cfr. *infra*).

c) Per ragioni puramente tecniche — considerando il fatto che per le basse frequenze l'indice HHI non può generare risultati statisticamente affidabili — sono state prese in esame soltanto parole con frequenza assoluta superiore a 200. È evidente che molte parole a collocazione ristretta che fanno parte di locuzioni idiomatiche o termini tecnici meno frequenti non hanno superato questa soglia (ad es., in italiano, *zozzo*, *crepapelle*, *bizzeffe*, *zuava*, *battibaleno* ecc.). Se da un lato, a causa della metodologia scelta, alcune delle PM meno frequenti mancano nella lista, dall'altro risulta chiaramente che molte PM costituiscono, con frequenze importanti, una parte integrante del lessico attuale e sono tutt'altro che marginali.

d) Come si è già accennato, le PM sono state finora studiate prevalentemente nell'ambito della fraseologia: i dati estratti dai corpora mostrano comunque che le PM, oltre alle locuzioni idiomatiche *stricto sensu*, fanno parte di numerosi tecnicismi, locuzioni avverbiali e preposizionali o costruzioni a verbo supporto.

3. CRITERI E IMITI DELL'ESTRAZIONE DELLE PM DAI CORPORA

Va notato che l'accezione delle PM può essere — dal punto di vista formale — ancora più *ampia*: L. Šíková Hrejsemnová, che si è dedicata al tema della monocollocabilità e alla possibilità di identificare le PM in base all'indice HHI e ad altri mezzi statistici nella sua tesi di laurea (2013, 16), distingue quattro tipi di monocollocabilità: verbi riflessivi, reggenze, locuzioni idiomatiche, polirematiche (multiword expressions).

Tutte queste categorie che dal punto di vista formale soddisfano ai criteri della monocollocabilità erano infatti presenti tra i risultati ottenuti dalla selezione automatica in base all'indice HHI. I dati raccolti sono stati sottoposti a un'ulteriore selezione manuale per eliminare parole nelle quali le restrizioni collocazionali erano dovute principalmente a ragioni grammaticali o lessico-grammaticali: sono stati esclusi i verbi riflessivi (*si sdraiò*), le reggenze (*proveniente da*), i participi dei tempi composti (*ha esordito*), i gerundi delle perifrasi (*sta succedendo*), i nomi modificati da sintagmi preposizionali (*una manciata di*) ecc.

La selezione automatica porta con sé anche un altro tipo di problema: il programma ha selezionato e ordinato i candidati PM basandosi sull'occorrenza dei loro collocati immediati (il primo collocato a sinistra e a destra della PM). Bisogna comunque tener presente che la frequenza dei rispettivi collocati è condizionata da tutta una serie di fattori: il collocato può essere soggetto a un'ellissi contestuale o lessicale o trovarsi in una posizione più lontana, fattore che non è stato preso in considerazione dal programma. Frequenti sono anche i casi di ellissi della testa nominale con conseguente uso sostantivato dell'aggettivo (*compact disc* → *il compact/*; *raggi infrarossi* → *gli infrarossi*).



Un'altra difficoltà risulta dalla natura stessa della formula HHI che indica con valori più alti le parole tendenti alla monoccollocabilità assoluta (*a malapena*) e le parole con un collocato dominante (*sistema immunitario*)³, mentre le PM con più collocati, situati più o meno allo stesso livello di frequenza (*midollo osseo/spinale/allungato; fino al midollo*), hanno valori molto più bassi. Per tutti questi motivi la sola selezione automatica non può bastare ed è comunque necessario un accurato triage manuale di tutti i candidati PM.

Un problema molto particolare è rappresentato dall'omonimia (o omografia) e dalla polisemia delle forme monoccollocabili. Il programma ha per esempio identificato la parola *capite* in base alla sua co-occorrenza con la preposizione latina *pro* (*pro capite*), tuttavia la frequenza assoluta di questa PM racchiude anche le occorrenze della forma flessa del verbo *capire*. Una situazione analoga può essere riscontrata nella parola *cubi*, che ha una collocabilità limitata solo in funzione di aggettivo (*metri cubi*). Raramente possono essere monoccollocabili ambedue gli omonimi o ambedue le accezioni della voce polisemica (cfr. *serial killer* e *serial televisivo/televisivi*).

La scelta di includere gli abbastanza frequenti casi di omonimia e polisemia può essere discutibile anche dal punto di vista metodologico: Veland per esempio inserisce nel suo corpus i casi di omonimia a livello di lemmi e esclude invece i casi di polisemia ("la polisemia implica per lo più distinzioni semantiche abbastanza sottili, non sufficienti comunque a dare origine a 'parole' diverse"; Veland 2006, 263). Dobrovolskij e Piirainen (1994, 456) escludono alcuni casi di "omofonia" e inglobano invece i "quasi-omonimi" (quando l'omonimo è strettamente legato a un linguaggio settoriale o gergo professionale e non è normalmente presente nel lessico mentale del parlante medio, quando si tratta di un arcaismo ecc.).

Come si è accennato sopra, i casi di omonimia e polisemia complicano anche il processo della selezione automatica, influenzando sul valore dell'indice HHI e alterando la proporzione tra la frequenza della PM e i suoi collocati. Nonostante tutte queste difficoltà gli autori hanno deciso di includerli nella lista (dopo aver verificato manualmente nel corpus la frequenza di queste PM nelle rispettive locuzioni), anche perché illustrano la tendenza attuale del lessico a preferire determinate accezioni (o più spesso una sola) di una parola omonima o polisemica ad altre.

4. SEMANTICA DELLE ESPRESSIONI MONOCOLLOCABILI: LOCUZIONI IDIOMATICHE E TERMINI TECNICI

Le parole monoccollocabili italiane con frequenza assoluta superiore a 200 estratte dal CORIS formano una collezione di 300 unità (per la lista completa delle PM con le frequenze delle PM e dei loro collocati cfr. Obstová 2016). La stragrande maggioranza

³ Tra i dati troviamo anche alcune parole che sembrano avere, almeno a prima vista, un raggio collocazionale abbastanza esteso (si tratta prevalentemente di aggettivi come *immunitario, infettive* o *igienica*). Queste *word-forms* rivelano soltanto una tendenza alla monoccollocabilità e vanno distinte dalle PM con un paradigma chiuso (nelle liste contrassegnate dal simbolo []).



delle PM forma con il suo collocato (i suoi collocati) una locuzione idiomatica in senso lato (131 PM, di cui 21 di origine straniera) o un tecnicismo (144 PM, di cui 40 forestierismi).⁴ 17 PM occorrono sia in locuzioni idiomatiche, sia in termini tecnici (*fino al midollo x midollo osseo/spinale*). Tra i forestierismi dominano gli anglicismi, quasi tutti ascrivibili alla categoria dei termini tecnici (*best seller, mountain bike, compact disc, exit poll, fast food, joint venture, know how* ecc.). Tra i latinismi individuati (15 in totale) la tendenza è inversa: più di due terzi formano locuzioni idiomatiche (*in loco, mea culpa, in extremis, aut aut*). Le sole due PM di origine francese fanno capo a due locuzioni idiomatiche (*tout court, tour de force*) e un termine tecnico di ambito culinario (*court bouillon*).

È interessante notare la leggera prevalenza dei termini tecnici⁵ rispetto alle espressioni idiomatiche, riscontrata nella fascia di frequenza superiore a 200; ciò mostra chiaramente che il fenomeno della monocollocabilità non si limita alla sola fraseologia (uno sguardo rapido alle PM meno frequenti sembra confermare ulteriormente questa tendenza). A tale riguardo vale anche l'affermazione di Čermák (Č. et al. 2016, 8) secondo cui la maggioranza delle espressioni idiomatiche presenta, dal punto di vista sintattico, carattere verbale o avverbiale (*mettere in risalto*), mentre i termini tecnici hanno carattere nominale (*anidride carbonica*).

5. STRUTTURE FORMALI

Le locuzioni che coinvolgono una PM — e che possono essere chiamate, in analogia con *cranberry expressions* (cfr. Trawiński et al. 2008), *espressioni monocollocabili* (EM) — presentano diverse strutture formali. Possiamo dividerle in locuzioni nominali, verbali, avverbiali, aggettivali, preposizionali, interiettive e pragmatiche: tutte queste categorie vengono qui descritte per ordine di importanza (numerosità). È evidente che alcune PM appartengono a due o più espressioni monocollocabili con strutture sintattiche diverse (cfr. *filato: zucchero filato/Tutto è filato liscio/andare dritto filato*). Il numero delle EM supera dunque quello delle PM.

Alcune PM, come abbiamo già menzionato *supra*, sono difficilmente classificabili;

4 Per locuzione idiomatica in senso lato intendiamo qui tutti quei casi che corrispondono alla definizione di Čermák (cfr. Čermák 2007, 83): “The idiom and phraseme is a unique combination of minimally two elements, one (or more) of which does not function in the same way in another combination (combinations), or it occurs in just one expression (or a severely limited number of such expressions)”, e dunque tutte quelle combinazioni nelle quali (almeno) uno dei costituenti non può subire una sostituzione paradigmatica (prova di commutazione) e che (nella maggioranza dei casi) denotano oggetti astratti (caratteristica che li distingue appunto dai termini tecnici). Consideriamo dunque espressioni idiomatiche anche alcune locuzioni avverbiali, preposizionali e talune costruzioni a verbo supporto (p. e. *alla rovescia, in bilico, alla stregua di, fare un clic*). Anche la categoria dei tecnicismi è intesa *lato sensu* (comprende polirematiche come *casa editrice* o *datore di lavoro*).

5 Bisogna tener presente che la ricerca si è basata su un corpus di riferimento (CORIS): è probabile che in un corpus specialistico, composto di soli testi scientifici, alcune delle PM qui identificate presenterebbero un raggio collocazionale meno ristretto.



in certi casi (ad es. *malapena, extravergine, sprovvista, impazzata, insaputa*) è quasi impossibile stabilire con precisione la parte del discorso della PM che non ha un'esistenza autonoma al di fuori della locuzione. Ciò nonostante cerchiamo di descriverle con le categorie grammaticali alle quali assomigliano di più (e con le quali vengono qualificate nei dizionari moderni, ad es. il GRADIT).

Le categorie grammaticali che rappresentano le parole monocollocabili vengono sottolineate, così come le PM nelle relative espressioni: N + A, *filo spinato*.

Nella classificazione a seguire vengono prese in esame soltanto le locuzioni italiane. Le locuzioni di origine straniera (che costituiscono 19,7% del totale) vengono trattate brevemente alla fine della sezione.

5.1. LOCUZIONI NOMINALI (38,8%)

Le locuzioni nominali rappresentano la struttura più frequente; la straordinaria numerosità delle polirematiche nominali è dovuta principalmente alla cospicua presenza di termini tecnici.

5.1.1. N + A (22,4%)

filo spinato; giorni feriali; acqua potabile; circolo vizioso

A questo tipo sono attribuibili, nonostante qualche perplessità, anche le locuzioni del tipo *casa editrice; erba cipollina; metri cubi ecc.*

5.1.2. N + PREP DI/A (+ DET) + N (4,7%)

fuoco d'artificio; principio di sussidiarietà; scala a chiocciola; campo di concentramento; fiore all'occhiello; sciopero/difesa ad oltranza

5.1.3. N + A (2,4%)

globuli rossi/bianchi; midollo osseo/spinale; inseminazione artificiale, cartoni animati

5.1.4. N + PREP + N (1,7%)

ciocca di capelli; datore di lavoro; dirittura d'arrivo; lasso di tempo

5.1.5. N + A (1%)

capro espiatorio; anidride carbonica

5.1.6. N + N; A + N (1,4%)

asili⁶ nido; stragrande maggioranza

5.1.7. ALTRI TIPI (1,4%)

io narrante; tran tran; vitto e alloggio; gratta e vinci; senno di poi

5.2. LOCUZIONI VERBALI (18%)

Le locuzioni verbali occupano il secondo posto nella scala della numerosità e costituiscono il 18 % delle EM identificate. Come è già stato osservato, la natura verbale ca-

6 Secondo i dati del corpus risulta monocollocabile solo al plurale.



ratterizza molte locuzioni idiomatiche. In questa categoria rientrano anche alcune costruzioni a verbo supporto. Va comunque notato che in taluni casi è difficile fare una distinzione fra le EM verbali e quelle avverbiali, soprattutto nei casi della monocollocabilità progressiva o “alla seconda”, come *gridare/cantare/urlare a squarciagola* dove la PM *squarciagola* dà origine alla locuzione avverbiale *a squarciagola* che a sua volta si combina con un numero molto limitato di verbi (appunto *gridare, cantare, urlare*); un altro esempio di questo tipo è la locuzione *svegliarsi di soprassalto*. In altri casi non è chiaro se il verbo vada considerato parte integrante della EM e se la costruzione vada interpretata come verbale o avverbiale (come [*essere*] in *voga*); inoltre la stessa locuzione può essere usata anche senza il verbo, in funzione di complemento di nome (*uno slogan in voga*), e dunque in funzione aggettivale.

5.2.1. V + PREP A/IN + (DET) + N (A) (8,5%)

andare a braccetto; andare in frantumi; finire nel mirino; mettere in palio; andare a passeggio; mettere a repentaglio; entrare in vigore; correre ai ripari; stare all'erta; essere in lizza; svegliarsi di soprassalto; cantare a squarciagola; cogliere alla sprovvista

5.2.2. V + (DET) + N (6,1%)

Questo tipo di struttura è caratteristico per delle costruzioni a verbo supporto: nella maggioranza dei casi, la PM costituisce la parte nominale dell'espressione e il verbo ha per lo più un significato molto ampio:

farsi beffe di q.; fare clic; fare le fusa; fare la spola; mettere la sordina; prendere/avere il sopravvento

5.2.3. V + A IN FUNZIONE PREDICATIVA (1,7%)

essere grato; essere stufo; essere lieto; stare zitto; passare inosservato

5.2.4. V + (DET) + N (1,4%)

In questa struttura la PM è la testa del SV. Il verbo (o una delle sue forme) monocollocabile è spesso semanticamente legato ai nomi che denotano una parte del corpo umano. *aggrottare la fronte; distogliere lo sguardo/il viso;⁷ rendersene conto*

5.2.5. ALTRI TIPI

avere da ridire

5.3. LOCUZIONI AVVERBIALI (13,6%)

Come è già stato notato, a volte è difficile fare la distinzione tra locuzioni avverbiali e verbali (cfr. 5.2.). Inoltre, in taluni casi la stessa PM può fare capo sia a una locuzione avverbiale, sia a una locuzione preposizionale (*nelle vicinanze x nelle vicinanze di; a ridosso x a ridosso di; alla stessa stregua x alla stregua di*). Le PM di questo tipo vengono incluse in ambedue le categorie.

⁷ I due verbi compaiono nella lista soltanto nella 3^a persona singolare del passato remoto (*aggrottò, distolse*); altre forme sono meno frequenti e non hanno superato la soglia delle 200 occorrenze nel corpus.



5.3.1. PREP + (DET) + N (10,2%)

Nella categoria delle locuzioni avverbiali, questa struttura ha la posizione predominante. La preposizione, semplice o articolata, è di solito *in*, *a*, raramente *di*. In queste locuzioni ci troviamo spesso di fronte a PM senza significato autonomo, difficilmente classificabili dal punto di vista grammaticale (cfr. *a malapena*, *a squarciagola*, *a malincuore*, *in auge* ecc.), anche se tali termini generalmente vengono considerati sostantivi (nel GRADIT, per esempio, *malapena* porta la qualifica di nome, *malincuore* di avverbio ecc.).

a casaccio; *in auge*; *in cattività*; *in subordine*; *di traverso*; *agli antipodi*, *nel/al contempo*; *nel frattempo*; *a malapena*; *a piacimento*; *a stento*; *in controtendenza*; *a dismisura*; *senza indugio*; *all'uopo*

5.3.2. PREP + (DET) + AVV (1,4%)

all'insù; *all'incirca*; *a malincuore*; *a tentoni*

5.3.3. PREP + (DET) + A (1%)

a ritroso; *di traverso*; *all'unisono*

5.3.4. ALTRI TIPI

in particolar modo; *pian piano*; *tutto sommato*; *senza batter ciglio*

5.4. LOCUZIONI AGGETTIVALI (4,4%)

È già stato menzionato (cfr. 5.2.) che talune EM possono avere carattere sia aggettivale, sia avverbiale (o verbale).

5.4.1. PREP + N (2%)

da capogiro, *di spicco*; *in voga*; *a tracolla*

5.4.2. ALTRI TIPI

nuovo di zecca; *territorialmente competente*; *naso all'insù*; *a tempo indeterminato*, *(senza) fine/scopo di lucro*

5.5. LOCUZIONI PREPOSIZIONALI (4%)

Le locuzioni preposizionali rappresentano solo il 4% di tutte le strutture presenti, bisogna comunque tener presente che il numero delle locuzioni non determina automaticamente la loro frequenza. Questo vale soprattutto per le locuzioni preposizionali che non di rado figurano tra le EM più frequenti (cfr. *nell'ambito di*, *in prossimità di*, *nelle vicinanze di*).

5.5.1. PREP + N + PREP (3,4%)

agli albori di; *nell'ambito di*; *sotto l'egida di*; *alla stregua di*; *nei paraggi di*, *nei pressi di*; *a suon di*, *all'insaputa di*

5.5.2. PREP + INF + PREP (0,6%)

a decorrere da; *a prescindere da*



5.6. LOCUZIONI INTERIETTIVE/LOCUZIONI PRAGMATICHE (1,4%)

bla bla; porca miseria; per intenderci

5.7. LOCUZIONI DI ORIGINE STRANIERA:

Tra i forestierismi che rappresentano il 19,7% del numero complessivo delle EM identificate prevalgono gli anglicismi (70%), i quali formano quasi esclusivamente locuzioni nominali di tipo N + N e A + N: *soap opera, tour operator, pole position, part time, serial killer; golden share/gol*. In alcuni casi in italiano sono monoccollocabili ambedue i costituenti (N + N, A + N): *task force; best seller; fast food, mountain bike*.

Importante è anche la presenza di locuzioni latine (26% tra i forestierismi) che hanno per lo più carattere avverbiale o aggettivale (*in primis, in extremis, in loco, a priori, ad personam, pro tempore, pro capite, in vitro*); meno frequenti sono le locuzioni nominali (*alter ego, nulla osta, status quo*).

Sono state identificate anche 3 locuzioni di origine francese, due nominali e una avverbiale (*tour de force; court bouillon; tout court*).

5.8. SINTESI

L'analisi delle strutture sintattiche rivela la netta prevalenza di locuzioni nominali, due volte più numerose rispetto alle locuzioni verbali che si piazzano al secondo posto, seguite da quelle avverbiali. Molto meno frequenti gli altri tipi — locuzioni aggettivali, preposizionali e interiettive. È curioso notare come queste percentuali corrispondano a quelle che Voghera (2004, 61) riporta per tutte le polirematiche presenti nel LIP (polirematiche nominali 39%, verbali 14,3%, avverbiali 14,3%, aggettivali 4,2%, interiettive 2,5%, congiunzionali 2,3%, preposizionali 2,2% ecc.). Questo confronto sembra dimostrare che le espressioni monoccollocabili italiane non differiscono affatto dal punto di vista strutturale dalle polirematiche “comuni”, formate da costituenti senza particolari limitazioni di collocabilità.

Se volessimo cercare di qualificare le singole PM italiane dal punto di vista grammaticale (nonostante le difficoltà legate proprio al carattere non autonomo di queste parole), arriveremmo ai seguenti risultati: prevalenza per i sostantivi (49%), seguiti da un numero sorprendentemente alto di aggettivi (40%) e a maggiore distanza da verbi (meno di 5%), avverbi e interiezioni. Se considerassimo anche le PM di origine straniera, gli anglicismi farebbero ulteriormente aumentare la categoria dei nomi, mentre le parole latine rafforzerebbero leggermente quella degli avverbi.

6. STRUTTURE COLLOCAZIONALI

Dalla tipologia sintattica e dagli esempi delle EM italiane riportati *supra* possiamo concludere che nella stragrande maggioranza dei casi (84%) il collocato si trova a sinistra della PM (qui sottolineata): questo vale per le locuzioni nominali (filo *spinato*, fuochi d'*artificio*), verbali (andare a *braccetto*, cogliere alla *sprovvista*), avverbiali (a *malapena*) e aggettivali (da *capogiro*). La struttura più frequente in termini di *collocational span* è PM — 1 (collocazione con una parola a sinistra della PM). Solo nel 10% dei casi il collocato si trova a destra, p. e. in alcune locuzioni nominali (*datore* di la-

voro, *stragrande* maggioranza) o in rarissime locuzioni verbali (*aggrottare* la fronte), mentre un residuo 6% delle PM richiede un collocato a entrambi i lati della base (si tratta esclusivamente di locuzioni preposizionali: *alla stregua di*, *agli albori di* ecc.).

Per l'italiano, lingua SVO e a testa iniziale, non è certo un risultato sorprendente considerando che il 91% delle PM individuate ha carattere di modificatore e solo il 9% di testa.

7. FREQUENZA

Il fatto che le PM rappresentino, dal punto di vista collocazionale, la periferia linguistica, non significa che esse siano periferiche anche sotto il profilo della frequenza. Tradizionalmente le PM vengono considerate marginali per il loro carattere spesso arcaico e opaco e anche la pratica lessicografica sembra confermare questa tendenza: infatti solo pochi tra i vocaboli a collocazione unica contrassegnati nei dizionari moderni dalla dicitura “solo nella loc.” e simili, analizzati da Veland (2006), superano le 200 occorrenze nel CORIS. La nostra ricerca si è basata invece sulle PM più correnti nel lessico attuale⁸: alcune delle parole, come *vigore* o *ambito*, superano la frequenza assoluta di 8000 o persino di 15000. Le parole monoccollocabili, come osserva Čermák (2014, 12), provengono generalmente da tutte le fasce frequenziali; è comunque interessante notare che l'occorrenza media delle PM italiane estratte dal CORIS è 611 e che la maggioranza delle parole (212 delle 300 PM) si colloca nella fascia tra 200 e 499 (di cui 104 PM nella fascia 200-299, 67 PM nella fascia 300-399 e 47 PM nella fascia 400-499). Con frequenze più alte il numero delle PM tende a diminuire progressivamente (65 PM uniformemente distribuite nella fascia 500-999; il limite di 1000 è stato superato solo da 17 PM).

8. “GRADO” DI MONOCOLLOCABILITÀ

La collocabilità (così come la monoccollocabilità) è un fenomeno scalare. L'indice HHI (vedi *supra*), in base al quale è stata effettuata la selezione automatica dal corpus, dovrebbe fornirci informazioni su quanto sia ristretto il raggio collocazionale delle singole parole. L'indice HHI può assumere valori da 0 a 1: più alto è il suo valore, più omogeneo è il contesto immediato delle parole esaminate. Da 1, raggiunto per l'italiano da una sola parola (*malapena*), i valori scendono rapidamente: quasi la metà delle 23 300 forme analizzate non raggiunge (né a destra, né a sinistra della PM) il valore di 0,1. Le PM dovrebbero essere contraddistinte da valori molto alti dell'indice, tuttavia considerata la natura scalare del fenomeno è impossibile stabilirne i contorni esatti. La selezione manuale è stata ulteriormente complicata dal fatto che la formula HHI indica con valori più alti le parole con un collocato unico o dominante mentre le

⁸ Come è già stato notato, per ragioni tecniche sono state prese in esame soltanto parole con frequenza assoluta superiore a 200. Per disegnare un quadro completo bisognerebbe naturalmente fare un'analisi dettagliata anche delle PM meno frequenti.



PM con più collocati, situati più o meno allo stesso livello di frequenza, hanno valori molto più bassi.

Consideriamo comunque utile riportare qui le PM con i valori più alti di HHI, vale a dire alcune delle forme caratterizzate dalla monocollocabilità pressoché assoluta.⁹ Tra parentesi è indicato se il valore massimo dell'indice è stato raggiunto a sinistra (L) o a destra (R) della PM:

HHI 1 — 0,90

malapena (L), *vitro* (L), *impazzata* (L), *malincuore* (L), *incirca* (L), *particolar* (R, L), *ca-saccio* (L), *primis* (L), *squarciagola* (L), *extremis* (L), *insù* (L), *quartier* (R), *tantino* (L) *galla* (L), *datori* (R), *stragrande* (R), *braccetto* (L), *repentaglio* (L), *ridosso* (L, R = 0,89), *suon* (L, R), *rendersene* (R), *datore* (R), *carbonica* (L), *rettorale* (L), *subordine* (L), *autonomia* (L), *ritroso* (L), *vedenti* (L), *priori* (L), *unisono* (L)

HHI 0,89 — 0,80

auge (L), *rado* (L), *tilt* (L), *bilico* (L), *sopravvento* (L), *spinato* (L), *sitter* (L), *disuso* (L), *seller* (L), *integrante* (L), *novero* (L), *lizza* (L), *concentramento* (L), *lasso* (R), *week* (R), *frat-tempo* (L), *ingrosso* (L), *espiatorio* (L), *sommato* (L), *intenderci* (L), *tentoni* (L), *stento* (L), *immunitario* (L), *egida* (L, R = 0,66), *operator* (L), *decorrere* (P, L), *prescindere* (L), *palio* (L), *dismisura* (L)

Nella stragrande maggioranza dei casi, il collocato delle PM con i valori dell'indice più alti è rappresentato da una preposizione (*a malapena*, *in vitro*). La struttura PREP + PM può a sua volta fare capo a una locuzione più ampia (*fecondazione in vitro*; *correre/battere all'impazzata*).

Colpisce anche la presenza di alcune forme apocopate (*particolar*, *quartier*, *suon*) che — almeno secondo i dati del corpus — non esistono nell'italiano attuale al di fuori delle rispettive locuzioni (*in particolar modo*, *quartier generale*, *a suon di qc.*).

9. CONCLUSIONI

Nelle pagine precedenti abbiamo cercato di descrivere e di analizzare le parole monocollocabili (e le rispettive locuzioni) più frequenti nell'italiano attuale estratte dal corpus di riferimento CORIS. Nonostante le PM vengano tradizionalmente studiate prevalentemente nell'ambito della fraseologia, i dati mostrano chiaramente che oltre a formare locuzioni idiomatiche le PM fanno capo a un numero non inferiore di termini tecnici. Non sorprende, nei tecnicismi, l'importante presenza di PM di origine inglese; tra i latinismi individuati la tendenza è inversa: 2/3 concorrono a formare locuzioni idiomatiche.

Le locuzioni che coinvolgono una PM presentano diverse strutture sintattiche. Dominano le locuzioni nominali, seguite da quelle verbali a avverbiali. Le locuzioni preposizionali, anche se meno importanti sotto il profilo della numerosità, sono spesso molto frequenti nel lessico attuale. La maggioranza delle espressioni idiomatiche presenta, dal punto di vista sintattico, carattere verbale o avverbiale mentre i termini tecnici hanno quasi esclusivamente carattere nominale.

⁹ Per la lista completa cfr. Obstová 2016.



Le PM sono per definizione difficilmente qualificabili dal punto di vista grammaticale; ciò nonostante si può concludere che la maggioranza di queste parole è ascrivibile alla categoria dei sostantivi e degli aggettivi, seguiti a maggiore distanza da verbi, avverbi e interiezioni. Colpisce l'alto numero di aggettivi (*spinato, potabile* ecc.) ai quali è stata finora dedicata scarsa attenzione nell'ambito della ricerca della collocabilità ristretta e che sembrano essere difficilmente identificabili con altri metodi di selezione.

La stragrande maggioranza delle PM italiane "sceglie" il suo collocato a sinistra e solo nel 10% dei casi il collocato si trova a destra. Questo risultato è in conformità con le caratteristiche tipologiche dell'italiano come lingua a testa iniziale, considerato che il 91% delle PM individuate ha carattere di modificatore e solo il 9% di testa.

Le PM, periferiche sotto il profilo collocazionale, vengono spesso considerate marginali anche dal punto di vista della frequenza. I risultati della ricerca mostrano che queste parole provengono da tutte le fasce frequenziali. Anche se la maggioranza delle PM identificate non supera, nel CORIS, la frequenza assoluta di 500, alcune di esse appartengono invece al vocabolario di alto uso e raggiungono frequenze importanti.

Le analisi in questo articolo non possono — nonostante la grande quantità dei dati esaminati — fornire che un contributo parziale alla problematica della collocabilità ristretta. Oltre al fatto che non sono state prese in considerazione le PM con frequenza inferiore a 200, è evidente che le liste sulle quali si è basata la nostra ricerca (sebbene esse offrano, per la prima volta, un campione sistematico e relativamente esauriente) non includono tutte le PM esistenti nell'italiano attuale: nessun corpus, per grande che sia, può rappresentare una lingua nella sua interezza. La monocollocabilità non è un fenomeno statico ma mutevole nel tempo e la nascita di nuove parole a collocazione estremamente ristretta non si limita al passato. L'identificazione di questi vocaboli nel lessico attuale potrebbe essere utile sia per la pratica lessicografica, che — a nostro giudizio — dovrebbe dedicare maggiore attenzione a questo fenomeno, sia per l'apprendimento dell'italiano come lingua straniera.

This study was written within the Programme for the Development of Fields of Study at Charles University, No. P10 Linguistics, sub-programme *Romance languages in the light of language corpora*.

RIFERIMENTI BIBLIOGRAFICI

- Aronoff, M. (1976) *Word Formation in Generative Grammar*. Massachusetts: The MIT Press.
- Bloomfield, L. (1933) *Language*. London: Allen Unwin.
- CoDII: *The Collection of Distributionally Idiosyncratic Items*. Accessibile a: <http://www.english-linguistics.de/codii/index.html>.
- Cvrček, V. (2013) *Kvantitativní analýza kontextu* [A Quantitative Analysis of Context]. Praha: Nakladatelství Lidové noviny.
- Čermák, F., J. Čermák, Z. Obstová e M. Vachková (2016) *Language Periphery. Monocollocable Words in English, German, Italian and Czech*. Amsterdam: John Benjamins.
- Čermák, F. (2014) *Periferie jazyka. Slovník monokolokabilních slov* [Language periphery. A dictionary of monocollocable words]. Praha: Nakladatelství Lidové noviny.



- Čermák, F. (2007) *Frazeologie a idiomatika česká a obecná* [Czech and General Phraseology]. Praha: Karolinum.
- Čermák, F. (1982) *Idiomatika a frazeologie češtiny* [Idiomatics and Phraseology of Czech]. Praha: Univerzita Karlova.
- De Mauro, T. (2000) *GRADIT. Grande dizionario italiano dell'uso*. Torino: UTET.
- Dobrovolskij, D. e E. Piirainen (1994) Sprachliche Unikalia im Deutschen. Zum Phänomen phraseologisch gebundener Formative. *Folia Linguistica*, 28/3-4, 449-473.
- Dobrovolskij, D. (1989) Formal gebundene phraseologische Konstituenten: Klassifikationsgrundlagen und theoretische Analyse. In: Fleischer, W., R. Große e G. Lerchner (eds) *Beiträge zur Erforschung der deutschen Sprache*, vol. 9, 57-78. Leipzig: Bibliographisches Institut.
- Fleischer, W. (1989) Deutsche Phraseologismen mit unikaler Komponente. Struktur und Funktion. In: G. Gréciano (ed.) *Europhras 88, Phraséologie Contrastive. Actes du Colloque International Kligenthal — Strasbourg, 12-16 mai 1988*, 117-126. Strasbourg: Université des Sciences Humaines, Département d'Etudes Allemandes, Collection Recherches Germaniques 1.
- Hausmann, F. J. (2004) Was sind eigentlich Kollokationen? Oder: wie pervers ist der wissenschaftliche Diskurs? In: Steyer, K. (ed.) *Wortverbindungen — mehr oder weniger fest*, 309-334. Berlin: De Gruyter.
- Hirschman, A. O. (1964) The paternity of an index. *The American Economic Review* 54/5, 761.
- Konecny, Ch. (2014) Unikale Lexeme — "Spuren" der Diachronie in der Synchronie? Eine Analyse anhand ausgewählter italienischer Phraseologismen. In: Melchior, L., A. Göschl, R. Rieger, M. Fischer e A. Voit (eds) *Spuren.Suche (in) der Romania. Beiträge zum XXVIII. Forum Junge Romanistik in Graz (18.-21. April 2012)*, 283-298. Frankfurt a. M.: Lang [Forum Junge Romanistik, 19].
- Obstová, Z. (2016) A frequency dictionary of Italian monocollocable words; An alphabetical dictionary of Italian monocollocable words. In: Čermák et al. *Language Periphery. Monocollocable Words in English, German, Italian and Czech*, 55-68; 89-97. Amsterdam: John Benjamins.
- Trawiński, B., M. Sailer, J. P. Soehn, L. Lemnitzer, e F. Richter (2008) Cranberry expressions in English and in German. *Proceedings of the LREC 2008 Workshop: Towards a Shared task for Multilingual Expressions (MWE 2008)*, 35-38.
- Šíková Hrejsemnová, L. (2013) *Monokolokabilní slova v češtině a dalších evropských jazycích* [Monocollocable words in Czech and other European languages], tesi di laurea non pubblicata. Praha: FF UK.
- Veland, R. (2006) Il concetto di collocazione unica e il valore di predizione della dicitura "solo nella loc." in uso nella pratica lessicografica. *Zeitschrift für Romanische Philologie* 122, 260-280.
- Veland, R. (2005) I vocaboli a collocazione unica nell'italiano di oggi. *Mémoires de la société néophilologique de Helsinki* 68, 331-339.
- Voghera, M. (2004) Polirematiche. In: M. Grossmann, e F. Rainer (eds) *La formazione delle parole in italiano*. Tübingen: Max Niemeyer Verlag.

CORPUS:

CORIS/CODIS, Università di Bologna. Accessibili a: http://corpora.dslo.unibo.it/coris_ita.html.

Zora Obstová

Department of Romance Studies, Faculty of Arts, Charles University in Prague
nám. J. Palacha 2, 116 38 Praha 1, Czech Republic
zora.obstova@ff.cuni.cz