

STUDI E SAGGI LINGUISTICI

LV (2) 2017

rivista fondata da

TRISTANO BOLELLI



Special Issue

WORD COMBINATIONS:
PHENOMENA, METHODS OF EXTRACTION, TOOLS

edited by

RAFFAELE SIMONE - VALENTINA PIUNNO

Edizioni ETS



Combinazioni di parole che costituiscono entrata. Rappresentazione lessicografica e aspetti lessicologici

RAFFAELE SIMONE - VALENTINA PIUNNO

ABSTRACT

Although the interest of literature in word combinations has significantly increased over the last decades, the full classification of their types and comprehensive collection of their forms is far from complete and flawless. This paper presents the results of a research project devoted precisely to the collection and analysis of Italian combinatorial lexicon. The investigation has both a theoretical and an empirical import. From the former point of view, it takes into account multiword units as specific forms of lexical entry. On the basis of the most recent achievements in lexicology, the notion of 'word format' is introduced and the vast variety of the combination formats is considered and classified. Various specific case studies of combinations are discussed. From the empirical point of view, the *CombiNet* dictionary is presented, an on-line lexicographic tool recording a sample of Italian multiword entries implementing an ample variety of entry formats. In addition to describing the lexicographic layout in detail, the paper illustrates the advanced research tool *CombiNet* is equipped with, as a device for both lexicographic and lexicological investigation.

KEYWORDS: word combinations, lexical entry, Italian lexicographic database, CombiNet, word format.

1. *Combinazioni di parole che costituiscono entrata*

Non è più necessario giustificare l'importanza di fenomeni come le combinazioni di parole, che ormai tutti ammettono senza difficoltà. Queste costituiscono un ambito molto vasto, in cui spiccano numerosi fatti significativi, poco conosciuti e meritevoli di indagine. Il presente articolo è dedicato all'analisi di una delle classi trascurate, cioè a quelle particolari *combinazioni di parole che costituiscono entrata* lessicale (o lemma)¹. Si tratta di combinazioni del tipo (1), ben distinte da quelle del tipo (2):

- (1) messa in piega, pronto soccorso, a briglia sciolta, per il verso giusto, andare avanti, prendere la palla al balzo
- (2) triste bilancio, bandire un concorso, aprire un conto

¹ Nel corso di questo articolo i termini *lemma* e *entrata* sono regolarmente usati come sinonimi.

Malgrado le somiglianze apparenti, la differenza tra le due categorie è netta: ciascuna di quelle in (1) costituisce un'entrata lessicale, cioè un'unità; per contro, ciascuno degli elementi in (2)², a dispetto della somiglianza superficiale con i precedenti, non costituisce entrata lessicale. La nostra indagine riguarda i tipi di lemma del primo tipo, cioè i *lemmi multiparola*. Il loro contrassegno sta nel fatto che hanno stabilità nel sistema e nella competenza e quindi devono essere considerati come *unità* in senso proprio.

Entrate di questo tipo sono di solito trascurate o poco riconosciute nella tradizione lessicografica, che ha cominciato a registrarle solo di recente. Ciò è sorprendente, perché esse presentano numerosi aspetti interessanti per l'analisi. Per indicarne alcuni, queste entrate sono numerosissime nel lessico e rispondono a pattern sintagmatici diversi. Inoltre possono essere osservate entro una pluralità di dimensioni, come quelle descrittiva, diacronica, tipologica, ecc., per ognuna delle quali offrono indicazioni rilevanti. Inoltre, costituiscono un problema computazionale, data la difficoltà di individuare i confini della combinazione.

Considerazioni di questo genere hanno spinto gli autori di questo articolo a progettare, insieme a colleghi di due università italiane, una ricerca pluriennale (chiamata *CombiNet*) sulle combinazioni di parole italiane che formano lemma³. L'indagine si inquadra nella cornice dell'attivissima ricerca internazionale sulla natura del lessico, sia dal punto di vista del significante che da quello del significato, e, in particolare, sui *formati di parola* (di cui diciamo qui sotto).

2. *Combinazioni di parole*

La nozione di 'formato di parola' è un comodo strumento per isolare e designare fenomeni che altrimenti non sarebbe facile catturare. In generale, un formato di parola è la struttura sintagmatica dell'entrata lessicale⁴.

² La lista in (2) contiene entità di diversa natura, che non interessa qui distinguere.

³ La ricerca *CombiNet* nasce da un progetto finanziato dal Ministero dell'Istruzione e della Ricerca (PRIN 2010-2011 n. 20105B3HE8), diretta dal primo autore di questo articolo, a cui hanno partecipato tre unità: una a Roma Tre (R. Simone, V. Piunno, L. Mereu, A. Pompei), una a Bologna (F. Masini e S. Castagnoli) e una a Pisa (A. Lenci e G. Lebani). Tra gli obiettivi della ricerca, l'identificazione, la raccolta e la catalogazione dei diversi tipi combinatori dell'italiano, nonché la rappresentazione delle proprietà combinatorie delle parole in un apposito database lessicografico. In particolare, quest'ultima parte della ricerca si è svolta presso le strutture del laboratorio TRIPLE di Roma Tre (<https://triplelab.wordpress.com/>).

⁴ In alcuni lavori di uno degli autori, esso è presentato come uno dei tipi di formato della cui stratificazione la parola è costituita. Vedi SIMONE (2017).

Questa nozione corrobora anche l'ipotesi che una lingua disponga di una molteplicità di formati di parola, ciascuno dei quali si manifesta in una molteplicità di lemmi. Inoltre, la nozione ha senso solo se si accetta l'ipotesi che le unità lessicali possano essere costituite anche da microsintagmi, cioè esser composte, oltre che da una sola parola (unità monoparola, quindi a struttura sintagmatica zero: *gatto, sedia, bontà, anello*, ecc.), anche da più parole formanti un'unità (unità multiparola: *messa in moto, faccia da schiaffi, prendere tempo*, ecc.). Il presente articolo si occupa solo dei formati multiparola. Siccome questi rientrano nel più ampio ambito delle combinazioni di parola, occorre dire qualcosa su queste prima di tornare al nostro tema specifico.

È evidente che le parole si combinano in sintagmi di varia estensione, complessità e struttura, e che gli elementi di questi sintagmi sono tenuti insieme da una Forza Coesiva⁵ di vario grado. I costituenti di alcuni sintagmi rimangono, per così dire, 'attaccati' dopo essere stati accostati nell'enunciazione, formando così combinazioni stabili nel sistema. I diversi tipi possibili di combinazioni sono stati catalogati e classificati in più occasioni e secondo più prospettive. Uno degli autori di questo articolo (Simone, 2006a; 2007), per esempio, ha proposto la seguente tassonomia:

- (3) Combinazioni
- a. *Combinazioni volatili*: meri sintagmi, fatti di parole che si aggregano nell'enunciazione e si dissolvono quando questa è esaurita;
 - b. *Combinazioni preferenziali*: aggregazioni di parole (coppie, triple, etc.) di varia natura sintattica, che *occorrono frequentemente insieme* ma non sono né collocazioni né costruzioni in senso proprio;
 - c. *Collocazioni*: aggregazioni di parole di varia natura, la cui peculiarità sta nel fatto che (a) incorporano sempre la propria testa, e (b) bloccano la sinonimia;
 - d. *Costruzioni in senso proprio* (rispondenti ad un insieme di requisiti specifici).

La lista di (3) può esser letta anche come una gerarchia (o una deriva diacronica), come in (4):

- (4) Combinazioni volatili > Combinazioni preferenziali > Collocazioni > Costruzioni in senso proprio

In altri termini, spostandosi da sinistra a destra della derivazione in (4) si incrementa una variabile importante, cioè la menzionata Forza Coesiva,

⁵ Per la nozione di Forza Coesiva, SIMONE (2007: 231 ss.).

il ‘legame’ che tiene insieme le parole. In questo senso la deriva da un grado all’altro di (4) può esser vista come una sorta di *grammaticalizzazione sintattica* (Lombardi Vallauri, 2004), forma peculiare del più conosciuto fenomeno della grammaticalizzazione: un sintagma originariamente volatile diventa *sempre più coeso* per codificare significati specifici.

Il grado di Forza Coesiva⁶ si misura secondo l’applicabilità di alcune restrizioni o trasformazioni⁷, tra cui in particolare⁸:

- (5) Restrizioni incorporate nella Forza Coesiva
- (a) Restrizioni morfologiche: è maggiore la Forza Coesiva delle combinazioni i cui costituenti perdono l’autonomia morfologica (per es., la possibilità di pluralizzazione: *piano di studio* vs. **piani di studi*; *acqua e sapone* vs. **acque e saponi*);
 - (b) Restrizioni sintattiche: perdita dell’autonomia sintattica dei costituenti (per es., possibilità di inversione dell’ordine: *essere di pessimo umore* vs. *essere di umore pessimo*; *treno ad alta velocità* vs. **treno a velocità alta*; per es., separabilità o ellissi dei costituenti: *letto a due piazze* vs. **letto a due grandi piazze*);
 - (c) Restrizioni semantiche (per es., grado di opacità semantica e conseguente perdita del significato compositazionale: *campo in erba* vs. *cantante in erba*);
 - (d) Restrizioni lessicali (per es., grado di *flagging*⁹ della combinazione: *dare del tu* vs. *dare del lei*);
 - (e) Restrizioni fonologiche (per es., perdita di autonomia accentuale dei costituenti, sandhi: *gratta e vinci* [grat:e’v:inʃi]) (Simone, 2007; 2008).

Le combinazioni che formano lemma sottostanno, in linea di principio, a tutte queste restrizioni. È difficile spiegare in modo soddisfacente il motivo per cui entrate lessicali diverse tendano a ‘saldarsi’ tra loro in modo stabile, ma il fatto come tale è indiscutibile: può darsi (come suggeriscono Bybee e Thompson, 1997 o Goldberg, 2006) che la frequenza di cooccorrenza abbia qualche ruolo nel convertire una combinazione volatile in una combinazione stabile.

⁶ Solitamente denominato alla francese *figement*, distinto in (a) *sintattico* (concerne la cristallizzazione sintattica dei costituenti del sintagma) e (b) *semantico* (concerne il grado di opacità semantica di una combinazione di parole). Le combinazioni *fisse* (o meglio *figées*) mostrano un alto grado di *figement* dal punto di vista sia sintattico sia semantico (PIUNNO, 2015).

⁷ Vedi ad esempio GROSS (1996), TUTIN e GROSSMANN (2002), LAMIROY (2003), VOGHERA (1994), DE MAURO (1999), VOGHERA (2004), SIMONE (2006), MASINI (2013), PIUNNO (2015).

⁸ Per una descrizione dettagliata delle diverse restrizioni si rimanda a PIUNNO (2015).

⁹ Il concetto di *flagging* è illustrato sotto nel testo.

2.1. *Livelli*

Le costruzioni si possono classificare in base a diversi criteri, sia formali che sostanziali. Qui ne saranno suggeriti alcuni¹⁰.

Quanto al livello sintagmatico, le costruzioni possono essere distinte come nello schema seguente:

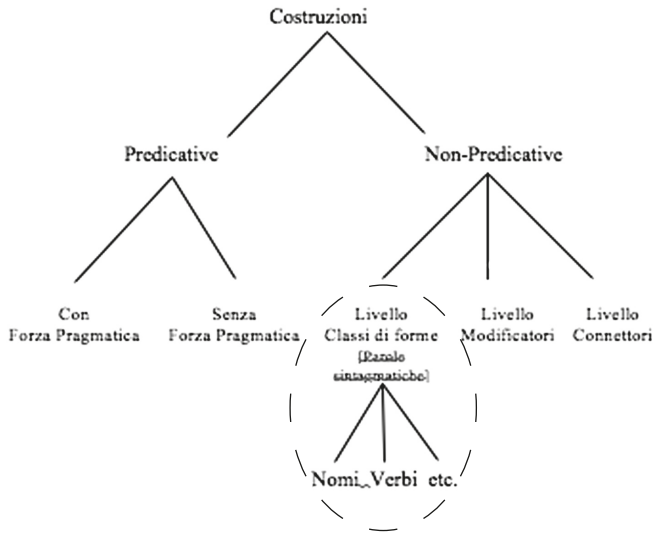


Figura 1. *Livelli di costruzioni.*

In questo schema la distinzione cruciale è quella tra Costruzioni Predicative e Non-Predicative, connessa con un'altra distinzione importante, quella tra Costruzioni Predicative dotate di Forza Pragmatica (FP) e Costruzioni Predicative prive di FP, della quale non possiamo dir nulla in questo luogo, rinviando a Simone (2006a; 2006b; 2007). Le Costruzioni Predicative incorporano una predicazione: possono essere composte da una sola clausola (come le costruzioni passive o le condizionali libere) o da due (come le causative); Costruzioni Non-Predicative sono tutte le altre. Ora l'argomento di questo articolo (e in generale della ricerca dalla quale esso deriva) è costituito dalle costruzioni non predicative che formano entrata, cioè da tutto ciò che rientra nel nodo della Figura 1 circondato da un tratteggiato.

¹⁰ Altre considerazioni sulla classificazione in ERMAN e WARREN (2000).

3. *CombiNet*

Passiamo ora alla presentazione di *CombiNet*. Illustreremo anzitutto i suoi presupposti e i suoi metodi, per poi concentrarci sulle combinazioni di parole che sono state estratte e su taluni loro aspetti. Nell'ambito di questo secondo obiettivo, illustreremo anche il dizionario che ne costituisce il risultato, e in particolare il layout lessicografico originale sviluppato per la rappresentazione dei dati.

L'indagine *CombiNet* si inserisce in uno degli ambiti di ricerca più produttivi dello studio attuale del lessico, quello dedicato all'analisi delle proprietà, dei criteri di classificazione e delle aree di sovrapposizione tra le categorie lessicali. In particolare, *CombiNet* risponde a una doppia esigenza, di natura lessicografica e lessicologica. Da un lato, mira a raccogliere e classificare le combinazioni e i formati di parola dell'italiano in base a principi teorici espliciti. Infatti, sebbene esistano dizionari specializzati di combinazioni italiane di parole¹¹, manca ancora un'opera che le raccolga e analizzi mediante criteri teoricamente espliciti.

Quanto all'analisi del lessico combinatorio, poi, *CombiNet* ha permesso di scoprire diversi nuovi formati nelle combinazioni, quasi del tutto ignorati in letteratura a dispetto del loro numero e della loro produttività. Dall'altro lato, *CombiNet* ha prodotto un esteso campione di un dizionario delle combinazioni di parole italiane, che costituisce, come si vedrà, uno strumento significativo anche a fini di ricerca.

4. *La raccolta dei dati*

4.1. *I corpora e l'estrazione dei dati*

CombiNet è basato interamente su corpora. Il database è costituito dalle combinazioni estratte dal corpus *La Repubblica*¹² (Baroni *et al.*, 2004), contenente testi di italiano giornalistico tratti dall'omonimo quotidiano, e dal corpus *Paisa*¹³, contenente materiale da Wikipedia e dal web.

La raccolta delle combinazioni è stata realizzata per mezzo di due

¹¹ Negli ultimi anni sono apparsi diversi dizionari di combinazioni italiane. Il più notevole è quello di LO CASCIO, edito in due versioni di diversa ampiezza (LO CASCIO, 2012; 2013); si veda anche URZI (2009).

¹² Cfr. <http://dev.sslmit.unibo.it/corpora/corpus.php?path=&name=Repubblica>.

¹³ Cfr. <http://www.corpusitaliano.it>.

software di estrazione, *EXTra* e *LexIt* (cfr. Lenci *et al.*, 2017; Castagnoli *et al.*, 2015; Castagnoli *et al.*, 2016). Il primo (Passaro e Lenci, 2016) è stato creato espressamente per estrarre combinazioni adoperando pattern combinatori predefiniti (sequenze di parti del discorso), caratterizzati da uno specifico ordine sintattico e un limitato numero di posizioni. Un insieme specifico di pattern è stato definito sulla base dei più recenti studi sulla combinatoria lessicale dell'italiano¹⁴. Sono state prodotte così liste di *frames* sintattici distinte secondo la funzione dell'unità combinatoria (per es. nominale, aggettivale, verbale, avverbiale, preposizionale). Il software permette di ottenere informazioni di diverso tipo sulle combinazioni: in particolare: la *log likelihood ratio* (l'indice statistico che misura il grado di combinatorietà tra più parole), la frequenza assoluta e le proprietà morfosintattiche (il femminile, il plurale, la presenza dell'articolo, ecc.).

Disponibile in rete, *LexIt* (Lenci, 2014; Lenci *et al.*, 2012), per parte sua, permette di raccogliere, mediante specifici indici statistici (per es., la *local mutual information*), le proprietà distribuzionali di nomi, verbi e aggettivi. Con *LexIt* si può estrarre la struttura argomentale dei lemmi sotto forma di *frames* sintattici e semantici (Lenci, 2014).

4.2. Parametri per la classificazione

Le combinazioni di parole di *CombiNet* sono state classificate secondo diversi criteri.

Il primo riguarda il *formato di parola*. Le diverse combinazioni estratte rappresentano una varietà di formati (chiamati *strutture* nella terminologia usata in *CombiNet*). Per es., le sequenze in (6), benché superficialmente diverse, sono rappresentate da una sola struttura o formato di parola: [NOME PREP NOME]; quelle in (7) da [PREP NOME]; quelle in (8) da [VERBO PREP NOME]:

- | | | |
|-----|---|---------------------|
| (6) | spazzolino da denti, carta di credito, chiavi in mano | |
| | | → [NOME PREP NOME] |
| (7) | alla moda, di getto, a quadri | → [PREP NOME] |
| (8) | chiamare per nome, andare a rotoli | → [VERBO PREP NOME] |

¹⁴ Per es., VOGHERA (1994; 2004), SIMONE e MASINI (2007), MASINI (2012), PIUNNO (2013; 2015; 2016).

CombiNet ha cominciato a costruire il repertorio dei formati di parola sottostanti alle combinazioni, ottenendo interessanti indizi sulla loro natura e la loro consistenza numerica. In particolare, tra i tratti distintivi di *CombiNet* sta il fatto di includere formati di parola ‘inediti’ (Piunno e Cominetti, in stampa), cioè poco conosciuti e ignorati (vedi § 7, più sotto).

Una volta costruiti i formati, vengono loro attribuite una o più categorie funzionali. Combinazioni con uguale struttura possono infatti svolgere funzione diversa secondo l’ambiente sintagmatico. Ad esempio le combinazioni in (9) e (10), pur avendo la stessa struttura, non svolgono la stessa funzione¹⁵. In altri casi, la funzione può essere multipla:

- | | | | |
|------|----|---------------------|--|
| (9) | a. | spazzolino da denti | [NOME PREP NOME] _{NOME} |
| | b. | chiavi in mano | [NOME PREP NOME] _{AGG AVV} |
| | c. | testa a testa | [NOME PREP NOME] _{AGG AVV NOME} |
| (10) | a. | alla moda | [PREP NOME] _{AGG AVV} |
| | b. | di getto | [PREP NOME] _{AVV} |
| | c. | a quadri | [PREP NOME] _{AGG} |

La combinazione in (9)a ha funzione di nome; quella in (9)b, di aggettivo o avverbio; in (9)c la sequenza *testa a testa* può svolgere più funzioni (aggettivo, avverbio o nome). Per parte sua, in (10)a la combinazione ha funzione doppia (aggettivo e avverbio), mentre le sequenze in (10)b-c svolgono una singola funzione (avverbio e aggettivo, rispettivamente).

In particolare, nella classificazione hanno un ruolo importante le restrizioni combinatorie. Le combinazioni differiscono infatti secondo il numero e la natura dei *flags* (Simone, 2006a; 2007), cioè il numero delle posizioni sintagmatiche riempite da elementi obbligati e il modo in cui sono riempite. Nei lavori appena citati sono distinte i seguenti gradi di *flagging*:

- (11) Gradi di flagging
- a. Basso: combinazioni costituite solo da categorie,
 - b. Medio: combinazioni fatte di categorie e di ‘pezzi’,
 - c. Alto: combinazioni fatte solo di ‘pezzi’.

Nella prima classe rientrano le combinazioni che contengono solo *categorie vuote*, che possono essere riempite in qualunque modo, a condizione

¹⁵ La funzione di una combinazione è segnalata in maiuscolotto sottoscritto subito dopo le parentesi quadre.

che siano rispettate alcune condizioni (per esempio, di natura semantica). La seconda classe include *patterns* contenenti sia categorie vuote (come sopra) sia elementi linguistici attualizzati. Un esempio può essere (12), la cui struttura è descritta in (13):

(12) per poco non sono caduto

(13) [*per poco non* [V {...}: fless]]

(13) si legge: la stringa *per poco non* è obbligatoria; è seguita da un verbo lessicale quale che sia, che dev'essere in forma flessa.

Il criterio del *flagging* si applica tanto alle costruzioni predicative (come nell'esempio precedente) quanto a quelle non predicative, e in particolare, per quel che ci riguarda qui, alle combinazioni che formano entrata lessicale. Questo è anzi uno dei punti di forza di *CombiNet*: il dizionario infatti non solo registra combinazioni di vario grado di lessicalizzazione, ma anche fenomeni combinatori con diverso grado di *flagging*. Include quindi sia formati di parola ad alto *flagging* (composti da lemmi fissi e non modificabili; li abbiamo chiamati *formati completamente riempiti*), sia formati a medio *flagging* (composti da lemmi fissi e insieme da posizioni 'vuote', che possono essere riempite variamente secondo specifiche restrizioni; li abbiamo chiamati *formati parzialmente riempiti*; vedi Piunno, in stampa).

I formati ad alto *flagging* (a) sono combinazioni stabili quanto ai lessemi che le compongono, (b) sintatticamente fisse e (c) non ammettono variazione (di natura lessicale o semantica). I formati parzialmente riempiti, invece, sono caratterizzati da un minore grado di coesione e di lessicalizzazione (e quindi da un minor grado di *flagging*) e hanno questa forma generale:

(14) Formati parzialmente riempiti¹⁶
 [*posizione fissa* + posizione variabile restrizione morfologica|RESTRIZIONE SEMANTICA]

Le combinazioni parzialmente riempite possono essere rappresentate da sintagmi con diversa funzione (per es., verbali o nominali) e estensione. Di seguito qualche esempio¹⁷:

¹⁶ In questa rappresentazione il corsivo indica le porzioni invariabili dal punto di vista lessicale e obbligatorie, mentre in pedice sono evidenziate le restrizioni morfologiche (in minuscolo) e le restrizioni semantiche (in maiuscolo) associate ad una porzione variabile del sintagma.

¹⁷ Ognuna delle voci delle tre liste che seguono avrebbe bisogno di specificazioni semantiche per quanto riguarda l'insieme dei componenti variabili (quelli in tondo). Tralasciamo qui questi dettagli per motivi di spazio.

- (15) [*prendere a* + NOME_{plurale}] = ‘colpire per mezzo di x’
prendere a pugni
prendere a calci
prendere a schiaffi
prendere a cannonate
prendere a sassate
prendere a bastonate
- (16) [NOME_{STRUMENTO} + *alla mano*] = ‘x pronto da usare’
 armi *alla mano*
 pistola *alla mano*
 documenti *alla mano*
 carte *alla mano*
 statistiche *alla mano*
 dati *alla mano*
- (17) [*avere* + DET + NOME + *facile*] = ‘essere incline a fare qualcosa connesso con x’
avere la lacrima *facile*
avere la pistola *facile*
avere la risata *facile*
avere il fischio *facile*
avere la querela *facile*

5. *Il layout lessicografico di CombiNet*¹⁸

Il dizionario *CombiNet* contiene circa 340 lemmi, costituiti da entrate nominali, aggettivali e verbali. Ci siamo limitati a queste, perché sono le categorie di uscita in cui ricadono le unità combinatorie complesse. Per la costruzione del lemmario sono state selezionate le parole incluse nella base di dati *Senso Comune*¹⁹, poi organizzate secondo la frequenza di occorrenza nel corpus *La Repubblica*. Secondariamente, tra le voci lessicali più frequenti sono state selezionate quelle con le più elevate capacità combinatorie.

Come nei dizionari convenzionali, ogni entrata registra le proprietà grammaticali (per es., la categoria grammaticale del lemma), con la sottoca-

¹⁸ Il layout lessicografico di *CombiNet* è stato ideato dall’unità di Roma Tre in collaborazione con le altre due unità di ricerca. Il software per la compilazione e la consultazione del database è stato creato da S. Dei Rossi, della società Websoup (Pisa).

¹⁹ Cfr. <http://www.sensocomune.it>.

tegorizzazione (per es., il genere e il numero nel caso dei nomi e la transitività o intransitività nel caso dei verbi). A seconda della ricchezza semantica del lemma, le voci sono articolate in accezioni e sottoaccezioni, ciascuna con una glossa contenente una breve definizione (Figura 2).



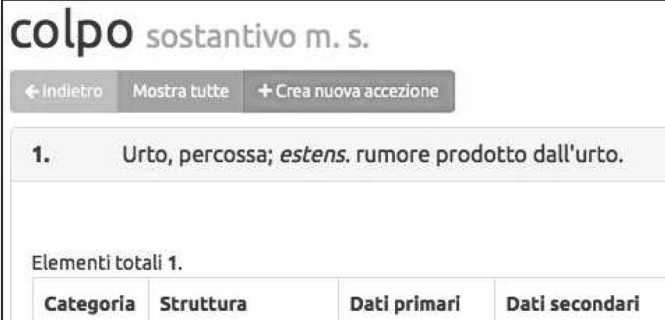
Figura 2. Esempio di entrata nominale di CombiNet.

Ogni accezione contiene una *matrice sintattica*, che rappresenta gli argomenti e gli aggiunti che sono tipicamente richiesti dal lemma in una specifica accezione. Ciascuna accezione contiene almeno una matrice sintattica.



Figura 3. Esempio di entrata verbale di CombiNet.

Per ogni lemma, e in particolare per ogni accezione, sono forniti i profili combinatori corrispondenti. In particolare, il layout lessicografico di *CombiNet* è articolato in quattro colonne: (a) *Categoria*, (b) *Struttura*, (c) *Dati Primari*, (d) *Dati Secondari* (Figura 4):



The screenshot shows a dictionary entry for the word "colpo" (sostantivo m. s.). At the top, there are navigation buttons: "← Indietro", "Mostra tutte", and "+ Crea nuova accezione". Below this, the first definition is listed: "1. Urto, percossa; *estens.* rumore prodotto dall'urto." Underneath the definition, it says "Elementi totali 1." and then a table with four columns: "Categoria", "Struttura", "Dati primari", and "Dati secondari".

Elementi totali 1.			
Categoria	Struttura	Dati primari	Dati secondari

Figura 4. Articolazione del layout lessicografico in quattro colonne.

Le prime due colonne sono dedicate alla rappresentazione delle proprietà sintattiche delle unità combinatorie. La prima è dedicata alla rappresentazione della categoria di uscita ('Categoria' in *CombiNet*, ovvero la funzione che una combinazione di parole assume nell'ambiente sintagmatico) o, in alcuni casi, una proprietà funzionale dell'unità combinatoria. *CombiNet* prevede dieci categorie: Nome, Aggettivo, Avverbio, Verbo, Verbo multiparola²⁰, Preposizione, Congiunzione, Binomi e trinomi, Espressioni idiomatiche, Interiezioni e Proverbi. Ad esse si aggiungono sequenze di soggetto e verbo (nell'ordine Nome_{SOGG} Verbo o Verbo Nome_{SOGG}).

La seconda colonna (denominata 'Struttura') registra il formato sintagmatico dell'unità combinatoria, espresso come struttura. Per esempio, la combinazione *carta di credito* sarà registrata con la struttura [NOME PREP NOME] e la categoria NOME. La tabella seguente riporta alcuni esempi di combinazioni di parole differenziate per struttura e categoria. Come si è osservato sopra, una stessa categoria può essere realizzata da più formati sintagmatici.

Le voci di *CombiNet* evidenziano in totale 323 formati sintagmatici diversi, distribuiti tra le diverse categorie di uscita. Nella seconda colonna del layout, infatti, appaiono numerosi tipi di entità, che danno l'idea della varietà di forme superficiali che un lemma può assumere, se appena si accetta l'idea che possano esistere lemmi multiparola. Man mano che l'analisi dei dati avanza, è possibile che la lista dei formati debba essere incrementata.

²⁰ Il termine *Verbo Multiparola* indica tutte le entrate verbali costituite da più parole costituenti un'unità inseparabile. I Verbi Multiparola sono di diversi sottotipi, dai verbi supporto (*fare benzina, mettere fretta*), a quelli con SP (*mettere a repentaglio, prendere per i fondelli*), ai verbi sintagmatici (*buttare giù, portare avanti, farsela addosso*).

CATEGORIA	STRUTTURA	ESEMPIO
NOME	NOME PREP NOME	<i>carta di credito</i>
	NOME PREP VERBO	<i>gomma da masticare</i>
	PREP NOME	<i>fuori programma</i>
AGGETTIVO	PREP NOME	<i>alla moda</i>
	AGG NOME	<i>biondo cenere</i>
	CONG NOME CONG NOME	<i>né carne né pesce</i>
AVVERBIO	PREP AGG NOME	<i>ad alta voce</i>
VERBO	VERBO _{sintagmatico} NOME	<i>mettere su casa</i>
MULTIPAROLA	VERBO PREP NOME	<i>chiamare in causa</i>
PREPOSIZIONE	PREP NOME PREP	<i>da parte di</i>

Tabella 1. *Articolazione dell'entrata lessicografica in categorie e strutture.*

È anche probabile che, immaginando idealmente di raccogliere tutte le combinazioni italiane che formano lemma, si ottengano numeri molto alti. Sin d'ora, infatti, l'analisi dei dati fornisce lunghe liste di strutture. In particolare, le categorie che contengono il maggior numero di formati sintagmatici sono le seguenti:

CATEGORIA	NUMERO DI STRUTTURE
VERBO MULTIPAROLA	87
AVVERBIO	71
AGGETTIVO	55
NOME	51
VERBO	44

Tabella 2. *Numero di formati di alcune categorie di CombiNet.*

Infine, la terza e la quarta colonna contengono gli esempi estratti dal corpus, ma si differenziano per il tipo combinatorio rappresentato, e in particolare per il grado di lessicalizzazione e la variabilità lessicale. *CombiNet* infatti registra, come si è detto, sia le combinazioni completamente riempite sia quelle parzialmente riempite. Tuttavia, i due tipi sono distinti topograficamente nel dizionario: ad ogni tipo è associata una specifica posizione nella voce. In particolare, la terza colonna ('Dati Primari') registra le combinazioni costituite da formati completamente riempiti. In questa colonna sono registrati diversi tipi sintagmatici, tra cui le collocazioni e le combinazioni preferenziali, le espressioni idiomatiche, le interiezioni e i proverbi, inseriti

come meri esempi combinatori. La colonna include anche polirematiche di diverso tipo (nominali, verbali, aggettivali, ecc.). Queste sono segnalate da un carattere differente (il sottolineato) e includono, nel caso di alta idiomatichità o significato opaco, una definizione semantica specifica. È il caso rappresentato nella Figura 5, per l'entrata *colpo*:

The screenshot shows the dictionary entry for 'colpo' (sostantivo m. s.). It includes navigation buttons: '← Indietro', 'Mostra tutte', and '+ Crea nuova accezione'. The first sense is '1. Urto, percossa; *estens.* rumore prodotto dall'urto.' Below this is a table with three columns: 'Categoria', 'Struttura', and 'Dati primari'.

Categoria	Struttura	Dati primari
Nome	~ Prep (Det) Nome	<ul style="list-style-type: none"> • ~ <u>di grazia</u> [mortale]. • ~ in testa. • ~ <u>in aria</u>, ~ <u>in canna</u>.

Figura 5. Esempi combinatori pienamente riempiti per l'entrata *colpo*.

La quarta colonna ('Dati Secondari') registra i formati di parola parzialmente riempiti, cui è dedicata una rappresentazione lessicografica peculiare. A ciascuno di essi è associato uno specifico formato di parola, come si vede dall'entrata *colpo*, in Figura 6:

The screenshot shows the dictionary entry for 'colpo' (sostantivo m. s.). It includes navigation buttons: '← Indietro', 'Mostra tutte', and '+ Crea nuova accezione'. The first sense is '1. Urto, percossa; *estens.* rumore prodotto dall'urto.' Below this is a table with four columns: 'Categoria', 'Struttura', 'Dati primari', and 'Dati secondari'.

Categoria	Struttura	Dati primari	Dati secondari
Nome	~ Prep (Det) Nome	<ul style="list-style-type: none"> • ~ <u>di grazia</u> [mortale]. • ~ in testa. • ~ <u>in aria</u>, ~ <u>in canna</u>. 	<ul style="list-style-type: none"> • [- <i>di</i> Nome_{arma}] ~ di artiglieria, ~ di bazooka, ~ di lupara, ~ di machete, ~ di manganello, ~ di rivoltella. • [- <i>di</i> Nome_{strumento}] ~ di ariete, ~ di bastone, maglio, ~ di martello, ~ di mazza, ~ di pennello

Figura 6. Combinazioni pienamente e parzialmente riempite per l'entrata *colpo*.

Mentre le combinazioni completamente riempite sono rappresentate in *CombiNet* come semplici liste di esempi, le combinazioni parzialmente ri-

empite sono rappresentate come liste di pattern sintagmatici. In particolare, ogni formato parzialmente riempito (per es., [*colpo di* NOME_{arma}] o [*colpo di* NOME_{strumento}]), come in Figura 6) include:

- (a) informazioni morfo-sintattiche (il numero e tipo di posizioni, la sequenza sintagmatica e, quando presenti, le restrizioni morfologiche, in pedice),
- (b) informazioni semantiche: restrizioni semantiche a cui ciascun formato è soggetto (testo in maiuscoletto in pedice),
- (c) il numero di posizioni riempite (in corsivo) e il numero di posizioni vuote (in tondo).

6. Funzioni di ricerca

Oltre che i dati lessicali finora descritti, il dizionario *CombiNet* contiene anche un sistema di ricerca integrato. Attraverso il database è infatti possibile eseguire numerosi tipi di ricerche: il software consente di raccogliere informazioni qualitative e quantitative sugli esempi e sui tipi combinatori inclusi.

6.1. Ricerche qualitative

Quanto alla ricerca ‘qualitativa’, riguardante i singoli lemmi e le loro proprietà combinatorie, *CombiNet* dispone di due funzioni: la ricerca semplice e la ricerca avanzata. Queste si distinguono, oltre che per il grado di complessità, anche per il tipo di indagine che permettono. La prima è una ricerca di tipo lessicografico convenzionale, la seconda di tipo lessicologico (consente di individuare le comunanze di tipo sintattico/semantico tra tutti i lemmi del dizionario, e i risultati che ne derivano sono utili per indagini lessicologiche).

La funzione di RICERCA SEMPLICE consente di ottenere informazioni di base relative sia all’entrata sia alle specifiche proprietà combinatorie.

Figura 7. Interfaccia per la funzione di ricerca semplice.

In particolare, la RICERCA SEMPLICE consente di:

- (a) consultare il dizionario ricercando un singolo lemma;
- (b) cercare i lemmi a partire dalle loro prime o ultime lettere (nella stringa di ricerca possono essere inseriti caratteri speciali, come l'asterisco, usato in sostituzione di uno o più caratteri);
- (c) cercare tutte le entrate appartenenti a una specifica categoria lessicale (per es., nominali, aggettivali, avverbiali) o a cui è stata attribuita una specifica categoria grammaticale (per es., nome femminile singolare, verbo transitivo, ecc.);
- (d) cercare le entrate omografe.

Le ricerche descritte si possono eseguire una per una; al contempo si possono richiedere più informazioni relative ad un lemma, per restringere il *range* di risultati possibili. La RICERCA SEMPLICE restituisce l'elenco dei lemmi che rispondono ai criteri selezionati.

L'altro tipo di ricerca possibile è la RICERCA AVANZATA. Si tratta di una funzione aggiuntiva che consente di effettuare ricerche più approfondite, e di usare *CombiNet* anche come strumento di analisi lessicologica. L'interfaccia RICERCA AVANZATA contiene tre sezioni: (a) lemma, (b) matrici sintattiche delle entrate verbali, (c) esempi combinatori.

Figura 8. *Interfaccia per la funzione di ricerca avanzata.*

Nella sezione dedicata al lemma è possibile effettuare ricerche più elementari: come con la RICERCA SEMPLICE, anche con la RICERCA AVANZATA è possibile selezionare un lemma specifico, una PoS o una qualsiasi informazione grammaticale associata all'entrata. In questa sezione c'è anche la possibilità di selezionare tutte le parole sintagmatiche e tutte le interiezioni o espressioni proverbiali inserite nel dizionario come esempio combinatorio. Si può inoltre preimpostare la modalità di visualizzazione dell'output della ricerca, selezionando i diversi campi da visualizzare (per es., lemma, accezione, categoria, struttura, ecc.).

La seconda sezione è dedicata alla visualizzazione delle matrici sintattiche rappresentata nelle diverse entrate verbali. In questo caso è possibile selezionare una specifica matrice sintattica ed ottenere la lista di tutti lemmi (e dei relativi esempi combinatori) in cui essa è presente.

L'ultima sezione della RICERCA AVANZATA è quella più innovativa, per le possibilità di indagine che apre. Tale sezione consente di esplorare tutte le proprietà combinatorie legate alle entrate e ha come output i dati combinatori. In questo caso, la RICERCA AVANZATA permette di selezionare una specifica categoria di output o una specifica struttura sintagmatica per ottenere liste di esempi combinatori estratti dai diversi lemmi. In questo caso la ricerca è in grado di estrarre combinazioni di parole strutturalmente e funzionalmente simili; l'output della ricerca contiene sia dati completamente riempiti

(‘dati primari’) sia quelli che lo sono solo parzialmente (‘dati secondari’). Infine, in questa sezione è possibile cercare una specifica parola (non necessariamente un’entrata lessicale) inserita in un qualsiasi esempio combinatorio (di tipo sia *primario* sia *secondario*).

Anche nel caso della RICERCA AVANZATA è possibile combinare assieme tutte le funzionalità di ricerca in una *query* multifunzionale. Ad esempio, *CombiNet* consente di rintracciare tutte le parole sintagmatiche che contengono un lemma dato, come nell’esempio seguente, che mostra combinazioni contenenti *mano*.

The screenshot shows the 'CombiNet' interface for an advanced search. The search bar contains the lemma 'mano'. The search is filtered by 'Entrata' (Entry) and 'Solo parole sintagmatiche' (Only syntagmatic words). The results are displayed in a list on the right side of the page, showing various combinations of the lemma 'mano' in different contexts, such as '(pl.) a piene mani [generosamente]', '(pl.) a mani basse [senza sforzo]', '(pl.) a mani nude [senza armi o senza guanti]', '(pl.) a mani vuote', '(pl.) allungare le mani', '(pl.) avere le mani bucate', '(pl.) avere le mani in pasta', '(pl.) dalle mani di [ad opera di]', '(pl.) finire nelle mani di qcn', '(pl.) in buone mani', and '(pl.) Mani Pulite'.

Figura 9. Esempio di una ricerca avanzata sul lemma *mano*.

Tra i risultati appaiono vari tipi di combinazioni, tra cui: nomi (per es., *Mani Pulite*), aggettivi (*in buone mani*), avverbi (*a piene mani*, *a mani nude*, *a mani vuote*), verbi (*allungare le mani*) e preposizioni (*nelle mani di*).

6.2. Ricerche quantitative

CombiNet contiene ulteriori pagine di RICERCA AVANZATA, che consentono di eseguire studi basati sui dati quantitativi. Le funzionalità di RICERCA AVANZATA che offrono informazioni quantitative sono differenziate in base all’oggetto di ricerca. Possono essere infatti riferite alle proprietà

grammaticali e combinatorie contenute nei lemmi, ai pattern combinatori (sia come categorie sia come strutture), alle combinazioni parzialmente riempite e alle matrici sintattiche.

Attraverso la pagina di RICERCA AVANZATA è possibile rintracciare due tipi di informazioni quantitative associate agli esempi combinatori, e in particolare (a) la frequenza assoluta di occorrenza della specifica combinazione nei corpora *La Repubblica* e *Paisà*, e (b) la *log likelihood ratio* calcolata mediante *ExTra* e *LexIt*.

The screenshot shows the 'CombiNet' interface. At the top, it says 'Visualizzo 1-20 di 43 elementi.' Below this is a section titled 'Singolo esempio'. A list of examples is shown, each with a tooltip providing statistical data. The tooltip for the example '(pl.) cerchia di amici' shows 'Indice statistico: 471.574.552.411' and 'Frequenza assoluta: 30'. Other examples include '(pl.) essere tra amici', '(pl.) circondarsi di amici', '(pl.) dagli amici mi guardi Iddio che dai nemici mi guardo io', '(pl.) rimanere amici', '(pl.) amici come prima', '(pl.) amici e parenti', 'da amico (consiglio da amico, prezzo da amico)', 'falso amico [parola che in due lingue è quasi omofona ma ha significati diversi]', and 'fare l'amico'.

Example	Statistical Index	Absolute Frequency
(pl.) essere tra amici	-	-
(pl.) cerchia di amici	471.574.552.411	30
(pl.) circondarsi di amici	-	-
(pl.) dagli amici mi guardi Iddio che dai nemici mi guardo io	-	-
(pl.) rimanere amici	-	-
(pl.) amici come prima	-	-
(pl.) amici e parenti	-	-
da amico (consiglio da amico, prezzo da amico)	-	-
falso amico [parola che in due lingue è quasi omofona ma ha significati diversi]	-	-
fare l'amico	-	-

Figura 10. *Indici statistici associati agli esempi combinatori.*

Un'altra pagina di RICERCA AVANZATA (la pagina 'Combinazioni') permette di estrarre e identificare i pattern combinatori più frequenti. In particolare, la pagina consente di visualizzare tutti i pattern che caratterizzano specifici tipi di entrate (per es., nominali, aggettivali, verbali) e che realizzano uno specifico tipo di categoria funzionale. I diversi formati sintattici possono essere ordinati per frequenza di occorrenza nel dizionario.

Di seguito sono registrati i pattern con funzione nominale più ricorrenti in *CombiNet*:

CombiNet Combinazioni POS Matrici sintattiche [-] Lemmario Q

Home / Combinazioni

Combinazioni

+ Crea Combinazione

Visualizzo 1-20 di 43 elementi.

Tipologia entrata	Categorie	Strutture	N° dati primari ↓↑	N° dati secondari
<input type="text" value=""/>	<input type="text" value="nome"/>	<input type="text" value=""/>	<input type="text" value=""/>	<input type="text" value=""/>
NOMINALE	Nome	~ Agg/Part. Pass.	2187	110
NOMINALE	Nome	Nome Prep (Det) ~	933	78
NOMINALE	Nome	~ Prep (Det) Nome	749	219
NOMINALE	Nome	~ Prep Nome	411	93
AGGETTIVALE	Nome	Nome ~	295	12
NOMINALE	Nome	Agg ~	254	15
NOMINALE	Nome	~ Nome	190	18

Figura 11. Pagina di ricerca avanzata 'Combinazioni': pattern nominali.

La figura mostra che i formati con funzione nominale (in maiuscoletto sottoscritto dopo le parentesi quadre) più frequenti sono:

- (18) Formati nominali di maggior frequenza
- d. $[\text{NOME} + \text{AGG/PART. PASS.}]_{\text{NOME}}$ acqua dolce, acqua distillata
 - e. $[\text{NOME} + \text{PREP} + \text{NOME}]_{\text{NOME}}$ bagaglio a mano, colpo di frusta

Nella figura che segue sono invece rappresentati i formati con funzione aggettivale, di cui i più frequenti sono:

- (19) Formati aggettivali di maggior frequenza
- a. $[\text{PREP} + \text{NOME}]_{\text{AGG}}$ a cuore, in piedi, da latte
 - b. $[\text{PREP} + \text{AGG} + \text{NOME}]_{\text{AGG}}$ di mezza età, di cattivo gusto

CombiNet Combinazioni POS Matrici sintattiche [-] Lemmario Q

Combinazioni

+ Crea Combinazione

Visualizzo 1-20 di 50 elementi.

Tipologia entrata	Categorie	Strutture	N° dati primari ¶	N° dati secondari
<input type="text" value=""/>	<input type="text" value="agg"/>	<input type="text" value=""/>	<input type="text" value=""/>	<input type="text" value=""/>
NOMINALE	Agg	Prep (Det) ~	186	0
NOMINALE	Agg	Prep (Det) Agg ~	78	11
NOMINALE	Agg	Prep (Det) ~ Agg	64	13
NOMINALE	Agg	Agg/Part.Pass. Prep (Det) ~	26	10
AGGETTIVALE	Agg	Prep (Det) ~ Nome	24	10
AGGETTIVALE	Agg	~ Prep Nome	22	7
NOMINALE	Agg	Prep (Det) ~ Prep (Det) Nome	20	12
NOMINALE	Agg	Prep Nome Prep ~	16	0

Figura 12. Pagina di ricerca avanzata 'Combinazioni': pattern aggettivali.

Infine, attraverso *CombiNet* è possibile raccogliere la lista dei pattern parzialmente riempiti (Figura 13) e delle matrici sintattiche delle entrate verbali (Figura 14).

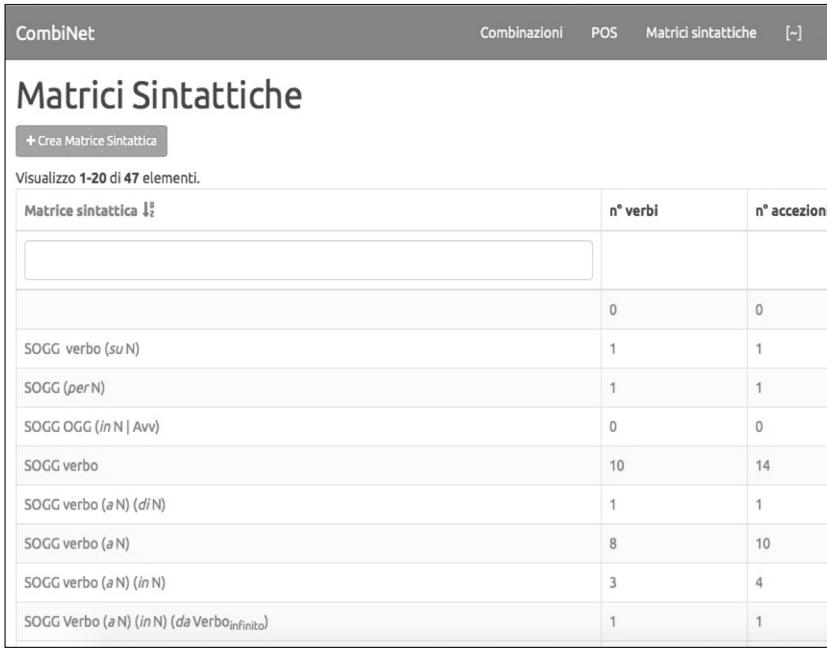
CombiNet Combinazioni Lemmario Q

Pattern parzialmente riempiti

Visualizzo 1-20 di 288 elementi.

Contenuto	Tilde Count ¶
<input type="text" value=""/>	
[NUM _{ord} ~]	4
[~ di N]	4
[~ di Det N]	4
[in ~ Agg]	4
[~ in N]	3
[per ~ Agg]	3
[in ~ Agg _{nazionalità}]	3
[~ N]	3
[~ per N]	3
[~ contro Det N]	2

Figura 13. Pagina di ricerca avanzata per le combinazioni parzialmente riempite.



CombiNet Combinazioni POS Matrici sintattiche [-]

Matrici Sintattiche

+ Crea Matrice Sintattica

Visualizzo 1-20 di 47 elementi.

Matrice sintattica	n° verbi	n° accezioni
<input type="text"/>		
	0	0
SOGG verbo (su N)	1	1
SOGG (per N)	1	1
SOGG OGG (in N Avv)	0	0
SOGG verbo	10	14
SOGG verbo (a N) (di N)	1	1
SOGG verbo (a N)	8	10
SOGG verbo (a N) (in N)	3	4
SOGG Verbo (a N) (in N) (da Verbo _{infinito})	1	1

Figura 14. Pagina di ricerca avanzata per le matrici sintattiche.

7. Prospettive di indagine

L'elaborazione del dizionario nelle sue diverse fasi ha fatto emergere numerosi fenomeni, alcuni dei quali non ancora analizzati o comunque trascurati. In conclusione di questo lavoro, vogliamo soffermarci su alcuni di questi: in particolare, sui formati con funzione di intensificatori (§ 7.1), i sintagmi verbali complessi con negazione obbligatoria (§ 7.2) e le combinazioni multifunzione (§ 7.3)²¹.

7.1. Formati di intensificatori

Il fenomeno degli intensificatori è abbastanza noto e non è questo il luogo per presentarlo nei suoi termini generali. Quel che interessa qui è che una parte notevole degli intensificatori dell'italiano sono costituiti da combina-

²¹ Si tratta di formati di parola caratterizzati da diversi gradi di lessicalizzazione, ma che meritano di essere inclusi tutti in un dizionario combinatorio: alcuni rappresentano semplici collocazioni (§ 7.1), altri rappresentano espressioni multiparola pienamente lessicalizzate (§ 7.2).

zioni, che ovviamente realizzano specifici formati. Tra questi, notiamo in particolare sei tipi, distinti per struttura sintagmatica, categoria di uscita e peculiarità semantiche.

Il primo tipo comprende alcuni nomi quantificatori che operano anche come intensificatori. Come è stato già notato (per es., Labov, 1984: 44), la classe degli intensificatori è spesso associata a quella dei quantificatori, il tratto in comune essendo la scalarità. I nomi quantificatori sono nomi ‘leggeri’ (Simone e Masini, 2014)²², cioè dalla semantica alleggerita, che occorrono come N_1 in strutture del tipo [NOME₁ PREP NOME₂] (per es., *bicchiere di vino*) e forniscono un’informazione qualitativa o quantitativa sull’intensione del N_2 , segnalando la classe semantica a cui esso appartiene (Simone e Masini, 2014: 58):

- (20) [DET NOME PREP]_{QUANTIFICATORE|CLASSIFICATORE}
 una *manciata* di (riso)
 una *briciola* di (pane)
 un *mazzo* di (fiori/chiavi)

Alcuni di questi possono essere usati anche come intensificatori (o, all’inverso, come mitigatori)²³, in specifiche sequenze, come in (21)-(22):

- (21) [DET NOME_{MASSA|COLLETTIVO} PREP]_{INTENSIFICATORE}
 un mondo di, un mare di, una manciata di
- (22) [DET NOME_{CONCRETO} PREP]_{INTENSIFICATORE}
 un sacco di, una montagna di

Il secondo tipo di intensificatori comprende aggettivi monoparola (23a, b) e multiparola (23c) (Piunno, 2013; 2015):

- (23) a. [NOME AGG]_{INTENSIFICATORE}
 freddo *gelido*, caldo *torrido*
- b. [AGG]_{INTENSIFICATORE} AGG_{sostantivizzato}
vero cretino, *totale* demente
- c. [NOME [PREP NOME]]_{INTENSIFICATORE}
 vestito *da sogno*, casa *da urlo*

²² Gli intensificatori di questa prima categoria sono una sottoclasse dei nomi classificatori, sottoclasse dei *light nouns* (SIMONE e MASINI, 2014). La loro funzione specifica consiste nel classificare semanticamente i lessemi con cui si accompagnano.

²³ È chiaro che il mitigatore non è che un intensificatore con segno rovesciato.

Il terzo tipo è rappresentato da SPrep di vario formato. Da un lato ci sono i SPrep con funzione aggettivale (o avverbiale), formanti espressioni iperboliche con significato rafforzativo o peggiorativo (Piunno, 2013; Benigni, in stampa):

- (24) a. [AGG [PREP NOME]]_{INTENSIFICATORE}
nuovo *di zecca*
b. [AGG [PREP VERBO_{infinito}]]_{INTENSIFICATORE}
bello *da morire*, pazzo *da legare*

Dall'altro ci sono le cosiddette *costruzioni aggettivali superlative* (Berlanda, 2013), costituite da coppie di aggettivi giustapposti e semanticamente legati da un rapporto di intensificazione:

- (25) [AGG_{INTENSIFICATO} AGG_{INTENSIFICATORE}]
ubriaco *perso*, stanco *morto*, bagnato *fradicio*

Nel quarto tipo rientrano verbi complessi funzionanti come intensificatori. Tra questi citiamo i seguenti, in cui tipicamente il verbo codifica l'iterazione di un atto, indicato dal SPrep che segue (con nome spesso eventivo):

- (26) [VERBO PREP NOME_{plurale}]
riempire di calci, *tempestare di domande*, *ubriacare di chiacchiere*

Il quinto tipo è rappresentato da SN con funzione avverbiale con la funzione di rendere scalare la proprietà o lo stato descritto dal verbo (Bolinger, 1972: 17):

- (27) [VERBO [DET NOME]]_{INTENSIFICATORE}
piacere *un sacco*, dispiacere *un casino*

7.2. Sintagmi verbali complessi con negazione obbligatoria

Una varietà di entrate sono formate da verbi complessi che occorrono con una negazione obbligatoria (*non*), nel senso che la medesima forma senza negazione o non è possibile o non ha il senso opposto. Si tratta di un fenomeno inedito, che non risulta osservato e che varrebbe la pena indagare anche comparativamente.

I formati con negazione obbligatoria sono diversi. Spiccano quelli verbali, spesso sintagmi discontinui con doppia negazione:

- (28) [NEG VERBO DET (AGG) NOME (AGG)]
non dare una lira (per), *non capire un fico secco* (di), *non vedere un bel niente*,
non capire un cavolo (di)²⁴

Queste combinazioni si distinguono per il SN (*una lira*, ecc.) che ha subito un processo di grammaticalizzazione fino a svolgere il ruolo di intensificatore di negazione. In tutti i casi la prima negazione è obbligatoria: la sequenza **vedere un fico secco* non è ammessa, almeno nel suo significato idiomatizzato.

A quest'ultimo formato si lega un fenomeno molto interessante nell'ambito del lessico polirematico e molto frequente: il formato rappresentato dal sintagma verbale complesso con negazione obbligatoria (per es., *non muovere un dito*) che esprime sempre una vera polarità negativa (per es., *non vedere l'ora*) e che non ha un corrispondente non negato (Piunno, in stampa).

In tutte le combinazioni menzionate la negazione non può essere omessa senza con ciò perdere il significato idiomatizzato. D'altro canto, tali combinazioni non sempre esprimono una vera polarità negativa. Possono quindi essere distinte in:

- (a) vere negative (per es., *non toccare cibo* vs. **toccare cibo*), il cui la negazione ha portata sul predicato;
- (b) false negative (per es., *non vedere l'ora*), in cui la negazione non ha una vera polarità negativa e l'insieme va trattato piuttosto come un *idiom*.

Come si mostra in Piunno (in stampa), più una negazione è 'falsa', maggiore sarà il grado di lessicalizzazione dell'intera sequenza. Di seguito sono riportati i pattern con negazione obbligatoria registrati in *CombiNet*. Si nota che il formato più frequente è quello con la doppia negazione:

²⁴ È ovvio che la seconda parte in corsivo, costituita da sintagmi apparentemente pieni, è in effetti un intensificatore di negazione. È per questo che l'insieme viene trattato da 'doppia negazione'.

Tipologia entrata	Categorie	Strutture	N° dati primari	N° dati secondari
		neg		
VERBALE	Verbo multiparola	Neg ~ (Det) Nome	9	0
NOMINALE	Verbo multiparola	Neg Verbo ~	7	0
VERBALE	Verbo multiparola	Neg ~ _{supporto} Nome	3	0
VERBALE	Verbo multiparola	Neg ~ Avv	2	0
VERBALE	Verbo	Neg ~ Avv	2	0
VERBALE	Verbo multiparola	Neg ~ Prep Avv	1	0
AGGETTIVALE	Verbo multiparola	Neg Verbo ~	1	0
NOMINALE	Verbo multiparola	Neg Verbo ~ Agg	1	0
NOMINALE	Verbo multiparola	Neg Verbo NUM ~	1	0
VERBALE	Verbo multiparola	Neg ~ _{supporto} Verbo ~ _{supporto}	0	5

Figura 15. *Formati di parola obbligatoriamente negati registrati in CombiNet.*

7.3. *Combinazioni multifunzione*

Vale infine la pena di menzionare le combinazioni con uguale formato sintagmatico ma con diversa funzione sintattica. Si tratta di combinazioni di parole che, a parità di forma, si differenziano solo per la funzione svolta nella frase e, per conseguenza, per le rispettive proprietà distribuzionali (Piunno, 2016). Il fenomeno è dovuto all'effetto di una trasposizione di livello, che permette a una forma di trasferirsi senza variazioni superficiali da una classe di parole a un'altra (Simone e Lombardi Vallauri, 2011: 132 ss.; Ramat e Jezek, 2009).

Ad esempio, la molte combinazioni rispondenti alla struttura [PREP NOME] possono essere usate come nome (per es., un originale *fuori onda*), aggettivo (per es., una trasmissione *fuori onda*) o avverbio (per es., esibirsi *fuori onda*). I pattern multifunzione non solo sono numerosi, ma sono anche rappresentati da formati di diverso tipo (Tabella 3)²⁵:

²⁵ Ai pattern multifunzione è tuttora riservata poca attenzione in letteratura (fatta eccezione per i *modificatori misti*, pattern preposizionali con duplice funzione aggettivale e avverbiale, cfr. PIUNNO, 2016).

CATEGORIA	FORMATI	ESEMPI
AGGETTIVALE O AVVERBIALE	NOME PREP NOME	<i>chiavi in mano</i>
	NOME PREP DET NOME	<i>mano nella mano</i>
	PREP NOME AGG	<i>a mani nude</i>
	PREP AGG NOME	<i>a buon mercato</i>
	PREP NOME PREP NOME	<i>a misura d'uomo</i>
AGGETTIVALE, AVVERBIALE O NOMINALE	PREP NOME	<i>fuori onda</i>

Tabella 3. *Pattern multifunzione.*

8. Conclusioni

Il progetto *CombiNet* ha prodotto un vasto repertorio di combinazioni di parole italiane e ha rivelato una gran massa di dati inediti, sconosciuti o malnoti. Ha inoltre creato un dizionario che registra e analizza le combinazioni di parole secondo criteri teorici espliciti. Il processo di realizzazione del dizionario di combinazioni ha mostrato che nell'individuare le categorie da includere in un dizionario non si può prescindere dai dati: solo lavorando su dati da corpora estesi si riesce a notare che i formati di parola di una lingua sono numerosi, comportano una sintagmatica ricca e variegata e possono essere molto diversi (in termini non solo strutturali, ma anche funzionali, semantici, ecc.). Inoltre, le funzionalità di ricerca incorporate nel dizionario hanno permesso di identificare fenomeni nuovi (per es., formati parzialmente riempiti, formati di intensificazione, negazioni obbligate).

Infine, da un punto di vista strettamente lessicale, l'estrazione dei dati e la loro classificazione indicano con chiarezza, se ancora ce ne fosse bisogno, che la nozione di entrata lessicale ha bisogno di una revisione approfondita (Simone, 2016): il lessico non è interamente monorematico, ma è caratterizzato da numerose combinazioni di parole che costituiscono entrata lessicale.

ABBREVIAZIONI

AGG	aggettivo
AVV	avverbio
DET	determinante
N	nome
NEG	negazione
PART. PASS.	participio passato
PREP	preposizione
SOGG	soggetto
SPrep	sintagma preposizionale

Bibliografia

- BARONI, M., BERNARDINI, S., COMASTRI, F., PICCIONI, L., VOLPI, A., ASTON, G. e MAZZOLENI, M. (2004), *Introducing the La Repubblica corpus: a large, annotated, TEI(XML)-compliant corpus of newspaper Italian*, in LINO, M.T., XAVIER, M.F., FERREIRA, F., COSTA, R. e SILVA, R. (2004, eds.), *Proceedings of the Fourth International Conference on Language Resources and Evaluation*, (LREC 2004, Lisbon, May 26th-28th 2004), ELRA - European Language Resources Association, Paris, pp. 1771-1774.
- BENIGNI, V. (in stampa), *Una festa da paura! Mi sono divertito da morire! Gli intensificatori iperbolici dell'italiano e la loro resa in russo*, in «Annales Universitatis Paedagogicae Cracoviensis Studia de Cultura».
- BERLANDA, S. (2013), *Constructional intensifying adjectives in Italian*, in KORDONI, V., RAMISCH, C. e VILLAVICENCIO, A. (2013, eds.), *Proceedings of the 9th Workshop on Multiword Expressions (MWE 2013, Atlanta, June 13th-14th 2013)*, Association for Computational Linguistics, Atlanta, pp. 132-137.
- BOLINGER, D. (1972), *Degree Words*, Mouton, The Hague.
- BYBEE, J. e THOMPSON, S. (1997), *Three frequency effects in syntax*, in JUGE, M.L. e MOXLEY, J.L. (1997, eds.), *Proceedings of the Twenty-Third Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on Pragmatics and Grammatical Structure*, Berkeley Linguistic Society, Berkeley, pp. 378-388.
- CASTAGNOLI, S., LEBANI, G.E., LENCI, A., MASINI, F., NISSIM, M. e PIUNNO, V. (2015), *Towards a corpus-based online dictionary of Italian word combinations: the CombiNet project*, Relazione presentata al COST ENeL WG3 meeting *Automatic Knowledge Acquisition for Lexicography*, Herstmonceux Castle, August 13th 2015 (http://www.elexicography.eu/wpcontent/uploads/2015/10/Paper_Castagnoli_etal_ENeL_final.pdf).

- CASTAGNOLI, S., LEBANI, G.E., LENCI, A., MASINI, F., NISSIM, M. e PASSARO, L.C. (2016), *POS-patterns or Syntax? Comparing methods for extracting Word Combinations*, in CORPAS PASTOR, G. (2016, ed.), *Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives (Full papers) - Fraseología computacional y basada en corpus: perspectivas monolingües y multilingües*, Tradulex, Geneva, pp. 101-114.
- DE MAURO, T. (1999, a cura di), *Grande Dizionario Italiano dell'Uso*. 6 voll., UTET, Torino.
- ERMAN, B. e WARREN, B. (2000), *The idiom principle and the open choice principle*, in «Text», 20, pp. 29-62.
- FILLMORE, C.J., KAY, P. e O'CONNOR, M.C. (1988), *Regularity and idiomaticity in grammatical constructions: the case of let alone*, in «Language», 64, pp. 501-538.
- GOLDBERG, A. (2006), *Constructions at work*, Oxford University Press, Oxford.
- GROSS, G. (1996), *Les expressions figées en français*, Ophrys, Paris.
- LABOV, W. (1984), *Intensity*, in SCHIFFRIN, D. (1984, ed.), *Meaning, Form, and Use in Context: Linguistic Applications*, Georgetown University Press, Washington D.C., pp. 43-70.
- LAMIROY, B. (2003), *Les notions linguistiques de figement et de contrainte*, in «Linguisticae Investigationes», 26, 1, pp. 1-14.
- LAZARD, G. (2001), *Études de linguistique générale: typologie grammaticale*, Peeters, Leuven-Paris.
- LENCI, A., LAPESA, G. e BONANSINGA, G. (2012), *LexIt: A Computational Resource on Italian argument structure*, in CALZOLARI, N., CHOUKRI, K., DECLERCK, T., DOĞAN, M.U., MAEGAARD, B., MARIANI, J., ODIJK, J. e PIPERIDIS, J.S. (2012, eds.), *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 2012 - Istanbul, May 23rd-25th 2012)*, European Language Resources Association (ELRA), Paris, pp. 3712-3718.
- LENCI, A. (2014), *Carving Verb Classes from Corpora*, in SIMONE, R. e MASINI, F. (2014, eds.), *Word classes: nature, typology and representations*, John Benjamins, Amsterdam-Philadelphia, pp. 17-36.
- LENCI, A., MASINI, F., NISSIM, M., CASTAGNOLI, S., LEBANI, G.E., PASSARO, L.C. e SENALDI, M.S.G. (2017), *How to harvest Word Combinations from corpora. Methods, evaluation and perspectives*, in «Studi e Saggi Linguistici», 55, 2, pp. 45-68.
- LO CASCIO, V. (2012), *Dizionario combinatorio compatto italiano*, John Benjamins, Amsterdam-Philadelphia.

- LO CASCIO, V. (2013), *Dizionario Combinatorio Italiano*, John Benjamins, Amsterdam-Philadelphia.
- LOMBARDI VALLAURI, E. (2004), *Grammaticalization of syntactic incompleteness: free conditionals in Italian and other languages*, in «SKY Journal of Linguistics», 17, pp. 189-215.
- MASINI, F. (2009), *Combinazioni di parole e parole sintagmatiche*, in MEREU, L. e LOMBARDI VALLAURI, E. (2009, a cura di), *Spazi linguistici. Studi in onore di Raffaele Simone*, Bulzoni, Roma, pp. 191-209.
- MASINI, F. (2013), *Parole sintagmatiche in italiano*, Caissa Italia, Cesena-Roma.
- PASSARO, L.C. e LENCI, A. (2016), *Extracting terms with EXTra*, Contributo presentato al Convegno *EUROPHRAS 2015 - Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives*, Málaga, June 29th - July 1st 2015.
- PIUNNO, V. (2013), *Modificatori sintagmatici con funzione aggettivale e avverbiale*, Tesi di dottorato inedita, Università Roma Tre, Roma.
- PIUNNO, V. (2015), *Sintagmi preposizionali come costruzioni aggettivali*, in «Studi e Saggi Linguistici», 53, 1, pp. 65-98.
- PIUNNO, V. (2016), *Multiword modifiers in Romance languages. Semantic formats and syntactic templates*, in COLSON, J.P. (2016, ed.), *Yearbook of Phraseology 7*, Mouton de Gruyter, Berlin-New York, pp. 3-37.
- PIUNNO, V. (in stampa), *Negated multiword expressions. Types, properties and degrees of lexicalization*, in FILATKINA, N. e STUMPF, S. (in stampa, eds.), *Konventionalisierung und Variation / Conventionalization and variation*, Peter Lang, Frankfurt a.M.
- PIUNNO, V. e COMINETTI, F. (in stampa), *Formati di parola inediti: dati emersi dal dizionario CombiNet*, in «Annales Universitatis Paedagogicae Cracoviensis Studia de Cultura».
- PIUNNO, V., MASINI, F. e CASTAGNOLI, S. (2013), *Studio comparativo dei dizionari combinatori dell'italiano e di altre lingue europee*. Rapporto tecnico inedito TRIPLE, Università degli Studi Roma Tre, Roma.
- RAMAT, P. e JEZEK, E. (2009). *On parts-of-speech transcategorization*, in «Folia Linguistica» 43, 2, pp. 391-416.
- SIMONE, R. (2006a), *Classi di costruzioni*, in GRANDI, N. e IANNÀCCARO, G. (2006, a cura di), *Zhì. Scritti in onore di Emanuele Banfi*, Caissa, Roma, pp. 383-409.
- SIMONE, R. (2006b), *Nominales sintagmáticos y no-sintagmáticos*, in DE MIGUEL, E., PALACIOS, A. e SERRADILLA, A. (2006, eds.), *Estructuras léxicas y estructuras del léxico*, Peter Lang, Berlin, pp. 221-241.

- SIMONE, R. (2007), *Constructions and categories in verbal and signed languages*, in PIZZUTO, E., PIETRANDREA, P. e SIMONE, R. (2007, eds.), *Verbal and Signed Languages. Comparing Structures, Constructs, and Methodologies*, Mouton de Gruyter, Berlin-New York, pp. 198-252.
- SIMONE, R. (2008), *Verbi sintagmatici come costruzione e come categoria*, in CINI, M. (2008, a cura di), *I verbi sintagmatici in italiano e nelle varietà dialettali. Stato dell'arte e prospettive di ricerca*, Peter Lang, Berlin, pp. 13-30.
- SIMONE, R. (2016), *Il dizionario del futuro*, in MARAZZINI, C. e MACONI, L. (2016, a cura di), *L'italiano elettronico. Vocabolari, corpora, archivi testuali e sonori*, Accademia della Crusca, Firenze, pp. 17-32.
- SIMONE, R. (2017), *Word as a stratification of formats*, in D'ALESSANDRO, R., IANNACCARO, G., PASSINO, D. e THORNTON, A. M. (2017, a cura di), *Di tutti i colori. Studi linguistici per Maria Grossmann* (edizione online), pp. 335-347.
- SIMONE, R. e LOMBARDI VALLAURI, E. (2011). *Natural constraints on language. The ergonomics of the software*, in «Cahiers Ferdinand de Saussure», 64, pp. 119-141.
- SIMONE, R. e MASINI, F. (2007), *Support nouns and verbal features: a case study from Italian*, in GREZKA, A. e MARTIN-BERTHET, F. (2007, eds.), *Verbes et classes sémantiques*, «Verbum», 29, pp. 143-172.
- SIMONE, R. e MASINI, F. (2014), *On Light Nouns*, in SIMONE, R. e MASINI, F. (2014, eds.), *Word classes: nature, typology and representations*, John Benjamins, Amsterdam-Philadelphia, pp. 51-73.
- TUTIN, A. e GROSSMANN, F. (2002), *Collocations régulières et irrégulières: esquisse de typologie du phénomène collocatif*, in «Revue française de Linguistique appliquée», 7, 1, pp. 7-25.
- URZÌ, F. (2009), *Dizionario delle combinazioni lessicali*, Convivium, Lussemburgo.
- VOGHERA, M. (1994), *Lessemi complessi: percorsi di lessicalizzazione a confronto*, in «Lingua e Stile», 29, 2, pp. 185-214.
- VOGHERA, M. (2004), *Polirematiche*, in GROSSMANN, M. e RAINER, F. (2004, a cura di), *La formazione delle parole in italiano*, Niemeyer, Tübingen, pp. 56-69.

RAFFAELE SIMONE

Dipartimento di Filosofia Comunicazione e Spettacolo

Università degli Studi Roma Tre

Via Ostiense 236

00146 Roma

raffaele.simone@uniroma3.it

VALENTINA PIUNNO

Dipartimento di Filosofia Comunicazione e Spettacolo

Università degli Studi Roma Tre

Via Ostiense 236

00146 Roma

valentina.piunno@uniroma3.it