

Valentina Piunno

Università degli Studi Roma Tre

Combinazioni di parole parzialmente riempite. Formati e rappresentazione lessicografica

1. Introduzione

Il presente contributo si colloca nell'ambito di un progetto di ricerca¹ dedicato allo studio delle combinazioni di parole della lingua italiana e alla loro rappresentazione in un'opera lessicografica, il dizionario *CombiNet*. *CombiNet* è un database lessicografico disponibile in rete² e basato su dati estratti da due corpora di italiano³, che raccoglie alcune combinazioni della lingua italiana. *CombiNet* è un importante strumento di analisi della combinatoria lessicale anche dal punto di vista lessicologico. Attraverso il database è infatti possibile effettuare numerosi tipi di indagini sulla struttura del lessico italiano: le ricerche avanzate disponibili in *CombiNet* consentono ad esempio di identificare le configurazioni sintagmatiche più produttive dell'italiano o di individuare nuovi formati di parola (cfr. Piunno 2016a, Simone – Piunno 2017).

Sulla scia del recente interesse rivolto alle proprietà combinatorie del lessico italiano, questo lavoro intende analizzare le proprietà sintattico-semantiche di un formato di parola specifico, quello rappresentato da strutture riempite solo parzialmente da unità lessicali. Il principale obiettivo del presente contributo consiste nell'identificare e analizzare alcuni formati parzialmente riempiti dell'italiano emersi durante la stesura del dizionario combinatorio, e di descriverne la relativa rappresentazione lessicografica scelta per *CombiNet*.

2. Combinazioni di parole: parametri di classificazione

I numerosi lavori dedicati dalla recente letteratura internazionale alle combinazioni di parole hanno impiegato diverse denominazioni e individuato diversi tipi

¹ Progetto PRIN 2010–2011 (n. 20105B3HE8): "CombiNet – Word Combinations in Italian: theoretical and descriptive analysis, computational models, lexicographic layout and creation of a dictionary" (Partenariato: Università Roma Tre, Università di Pisa, Università di Bologna).

² Sito internet (con accesso temporaneamente limitato): www.combinet.fileli.unipi.it.

³ Il corpus *La Repubblica* (<http://dev.sslmit.unibo.it/corpora/corpus.php?path=&name=Repubblica>) e il corpus *Paisà* (<http://www.corpusitaliano.it>).

classificazioni, delineando un panorama estremamente complesso (Piunno 2015). Le combinazioni di parole sono tradizionalmente classificate sulla base delle restrizioni a cui sono soggette. Uno dei criteri più frequentemente adoperati per la distinzione delle unità combinatorie è rappresentato dal grado di *figement*⁴ (o *forza coesiva*, cfr. Simone, 2007) e di lessicalizzazione, solitamente misurato sulla base dell'applicabilità delle seguenti restrizioni e delle relative trasformazioni⁵:

- Restrizioni morfologiche (es. pluralizzazione e perdita dell' autonomia morfologica dei costituenti);
- Restrizioni sintattiche (es. inversione dell'ordine dei costituenti, separabilità dei costituenti, ellissi di un costituente);
- Restrizioni semantiche (es. idiomaticità o opacità semantica, e conseguente perdita del significato compositivo);
- Restrizioni lessicali (es. variabilità lessicale o presenza di arcaismi)
- Restrizioni fonologiche (es. perdita di autonomia accentuale)

Un ulteriore criterio usato in letteratura è quello rappresentato dal *formato sintagmatico* della combinazione, vale a dire dalla configurazione sintagmatica alla base di un'unità combinatoria: es. il nome sintagmatico *punto di vista* è rappresentato dalla configurazione [Nome Preposizione Nome]⁶, mentre il verbo complesso *mandare in onda* dalla configurazione [Verbo Preposizione Nome]. A questo si lega la funzione che l'unità combinatoria svolge (definita in *CombiNet* come *categoria di uscita* di una combinazione), che rappresentiamo in pedice negli esempi che seguono:

- (1) *punto di vista* [Nome Preposizione Nome]_{NOME}
 (2) *mandare in onda* [Verbo Preposizione Nome]_{VERBO}
 (3) *chiavi in mano* [Nome Preposizione Nome]_{AGG|AVV}
 (4) *faccia a faccia* [Nome Preposizione Nome]_{AGG|AVV|NOME}

Come si evince dagli esempi in (3)–(4) una stessa unità combinatoria può avere più categorie di uscita e, per conseguenza, può svolgere molteplici funzioni all'interno della frase: la sequenza *faccia a faccia* può essere usata in funzione aggettivale (es. *un dibattito faccia a faccia*), avverbiale (es. *discutere faccia a faccia*) o nominale (es. *un faccia a faccia televisivo*).

⁴ Il *figement* consiste nella cristallizzazione dell'unità combinatoria. Il *figement* viene tradizionalmente distinto in i) *sintattico* (relativo alla cristallizzazione dei costituenti del sintagma) e ii) *semantico* (relativo all'opacità semantica dell'unità combinatoria). Le combinazioni di parole sono definite *fisse* (o meglio, *figées*) se mostrano un *figement* di tipo sia sintattico sia semantico (Piunno 2015).

⁵ Gross (1996), Mejri (1997), Tutin-Grossmann (2002), Lamiroy (2003), Voghera (1994), De Mauro (1999), Voghera (2004), Simone (2006), Masini (2013), Piunno (2013, 2015).

⁶ Nell'ambito del presente lavoro, sarà perlopiù utilizzato il termine "formato sintagmatico" per riferirci al pattern sintagmatico che caratterizza una combinazione. I termini "configurazione sintagmatica" e "struttura" (terminologia usata nel database *CombiNet*) verranno utilizzati come sinonimi, e per conseguenza saranno alternati al primo.

Per la classificazione delle unità combinatorie inserite in *CombiNet* sono stati considerati, in aggiunta a quelli appena menzionati, altri parametri di classificazione.

In primo luogo, facendo riferimento ai principi della *Grammatica di Categorie e Costruzioni* (Simone 2006, 2007), ci siamo avvalsi della nozione di *formato di parola*, vale a dire una struttura linguistica in grado di incorporare e rappresentare le informazioni semantiche (i *formati semantici*) e le strutture sintagmatiche (i *formati sintagmatici*) associate ad una parola. Il formato di parola si caratterizza dunque per la presenza di più componenti, ed è rappresentabile secondo la seguente struttura:

(5) *Formato di parola:*

Formato sintagmatico/semantico: SEQUENZA DI PoS_[CLASSIFICAZIONE ONTOLOGICA]
 Semantica della combinazione: _____

I formati semantici, da un lato, costituiscono dei pattern in grado di organizzare il significato di una parola, operando a un livello concettuale-cognitivo. I formati sintagmatici, dall'altro, costituiscono le strutture linguistiche specifiche a cui i formati semantici sono collegati. Combinazioni di parole che derivano dagli stessi formati condividono necessariamente delle proprietà morfo-sintattiche e semantiche, dando vita a gruppi combinatori semanticamente e sintatticamente affini (cfr. "insiemi sintetici", Sechehaye 1950[1926]; "serie (para)sinonimiche", Blanco 2001).

In secondo luogo, abbiamo preso in considerazione il grado di *variabilità paradigmatica* (*flagging*, cfr. Simone – Piuanno 2017) associato a gruppi di combinazione di parole. Durante la fase di compilazione delle voci del dizionario *CombiNet*, infatti, si sono distinti per quantità e variabilità di struttura dei gruppi di combinazioni di parole apparentemente lessicalizzate e coese, ma caratterizzate da una tendenza alla variabilità lessicale (Piuanno 2013). Si osservino gli esempi seguenti:

- (6) di {*prima* | *seconda*} mano
 (7) di {*alta* | *bassa* | *ottima* | *pessima*} qualità

La variabilità paradigmatica i) coinvolge uno o più costituenti, ii) è spesso soggetta a restrizioni semantiche specifiche, iii) seleziona lessemi semanticamente affini, iv) può essere rappresentata da *range* paradigmatici più o meno estesi.

In terzo luogo, la classificazione delle unità combinatorie è stata effettuata tenendo in considerazione il *grado di schematicità* delle sequenze, vale a dire il loro livello "specificità lessicale" (cfr. Fillmore et al. 1988). Combinazioni che condividono uno stesso formato di parola, e che sono sintatticamente coese possono essere variabilmente specificate a livello lessicale. I pattern sintattico-semantiche (formati di parola) possono essere rappresentati schematicamente attraverso configurazioni più "astratte", dal livello di produttività variabile. Il grado di produttività di un formato di parola muta al variare del livello di schematicità (Bybee 2010): inoltre, più la sequenza è specificata lessicalmente, più sarà probabile che il formato di parola sia produttivo. Inoltre, un basso livello di specificità lessicale è spesso associato a una alta predicibilità semantica del formato di parola.

3. Le combinazioni di parole parzialmente riempite

I processi di raccolta e di analisi dei dati per la stesura del dizionario *CombiNet*, avvenuti per mezzo di due diverse tecnologie di estrazione (cfr. Lenci *et al.* 2012; Passaro & Lenci 2016; Castagnoli *et al.* 2015, 2016; Piunno 2016b), hanno stimolato alcune nuove riflessioni sul concetto di *combinazione di parole*, evidenziando una moltitudine di fenomeni combinatori del lessico italiano (es. collocazioni, parole sintagmatiche, espressioni idiomatiche) che sono spesso riconducibili ad una semantica stabile e disciplinati da gruppi ristretti di co-selezioni. Tra le varie combinazioni sono emersi in particolare due macro tipi combinatori: a) le combinazioni di parole che sono interamente riempite da unità lessicali e b) quelle che lo sono solo parzialmente.

Le prime sono combinazioni stabili dal punto di vista delle proprietà lessicali e morfosintattiche, non sono soggette a variazioni di alcun tipo e sono rappresentative del più alto grado di cristallizzazione (per tale motivo perdono spesso il significato compositazionale per acquisire significati metaforici). Costituiscono i cosiddetti *formati completamente riempiti*, caratterizzati da un pattern semanticamente e sintatticamente fisso:

(8) parole sintagmatiche	(<i>punto di vista; mandare in onda</i>)
verbi supporto	(<i>fare benzina</i>)
espressioni idiomatiche	(<i>prendere il toro per le corna</i>)
interiezioni	(<i>grazie al cielo!</i>)
proverbi	(<i>occhio per occhio, dente per dente</i>)

Le seconde sono rappresentate da gruppi combinatori che i) condividono la medesima configurazione sintagmatica e ii) specifiche proprietà semantiche, e iii) si caratterizzano per la presenza di posizioni vuote, che possono essere riempite da materiale lessicale semanticamente omogeneo. Queste ultime sono quindi aperte alla variazione lessicale e presentano una caratterizzazione semantica stabile. Le combinazioni parzialmente riempite si configurano attraverso sintagmi di diversa natura (es. nominale, verbale, avverbiale, ecc.), che possono essere rappresentati secondo un pattern sintattico-semantico che ne definisce contestualmente la struttura sintagmatica e le restrizioni semantiche. I pattern sintattico-semantici possono veicolare diversi tipi di significato e possono rappresentare livelli di produttività variabili: nuovi esempi combinatori possono essere generati dai formati di parola esistenti, e convenzionalizzarsi per mezzo della ripetizione e, quindi, dell'alta frequenza d'uso (Bybee 2010). Le combinazioni parzialmente riempite contengono alcune posizioni fisse e altre posizioni variabili, sottoposte spesso a specifiche restrizioni semantiche e pertanto saturabili mediante materiale lessicale semanticamente omogeneo (Bybee 2010):

(9) [*posizione fissa* + **posizione variabile** RESTRIZIONE SEMANTICA]⁷

⁷ Nella rappresentazione qui fornita, il corsivo serve a indicare lo slot invariabile dal punto di vista lessicale, mentre il grassetto evidenzia lo slot soggetto a variazione. In pedice sono indicate le restrizioni semantiche associate alla porzione variabile.

Ad esempio, quando il verbo *dare* è seguito dalla preposizione *per* e da un aggettivo o un participio passato (solitamente, ma non sempre, connotati negativamente, es. *vinto, morto, spacciato*), assume lo specifico significato di “considerare”:

- (10) Formato sintagmatico/semantico: [*dare per* + Agg/Part. Pass]
 Semantica della combinazione: *considerare come X_{AGG}*
 Esempi: dare per **vinto**, dare per **favorito**, dare per **morto**, dare per **buono**, dare per **spacciato**, dare per **assodato**, dare per **scomparso**, dare per **vincente**, dare per **sicuro**

Riassumendo, i formati di parola parzialmente riempiti: i) si configurano attraverso un formato superficiale comune, ii) presentano “caselle vuote” che possono essere occupate da materiale lessicale semanticamente omogeneo, iii) sono sottoposte a restrizioni semantiche, iv) condividono specifiche proprietà semantiche, v) occorrono alternativamente nello stesso contesto sintattico.

3.1. Tipi di combinazioni parzialmente riempite

Sulla base di quanto specificato nei paragrafi precedenti, le combinazioni parzialmente possono essere distinte secondo: i) la categoria di uscita della combinazione, ii) la semantica, iii) la fissità della struttura (il livello di variabilità paradigmatica, l'estensione del *range* paradigmatico e, per conseguenza, per il livello di produttività della struttura).

3.1.1. Categoria di uscita

Dal punto di vista della categoria di uscita, le combinazioni di parole parzialmente riempite possono essere di diverso tipo, come ad esempio nominali (11), verbali (12), aggettivali (13):

(11) Combinazioni nominali

- Formato sintagmatico/semantico: [*colpo di* + Nome_{ARMA}]
 Semantica della combinazione: *esplosione provocata dalla messa in funzione di un'arma da fuoco*
 Esempi: colpo di **pistola**, colpo di **fucile**, colpo di **lupara**, colpo di **cannone**, colpo di **mitra**

(12) Combinazioni verbali

- Formato sintagmatico /semantico: [*dare del* + Agg/Part. Pass._{connotazione negativa}]
 Semantica della combinazione: *attribuire l'appellativo di X_{AGG}*
 Esempi: dare del **cretino**, dare del **bugiardo**, dare del **visionario**, dare dell'**idiota**, dare del **teppista**, dare del **fallito**, dare del **mafioso**, dare del **buffone**

(13) Combinazioni aggettivali

- Formato sintagmatico/semantico: [*facile a* + Nome]
 Semantica della combinazione: *essere incline a X_{NOME}*

Esempi: facile all'**entusiasmo**, facile alla **commozione**, facile all'**insulto**, facile alla **polemica**, facile alla **riッサ**, facile all'**ira**

3.1.2. Semantica

Dal punto di vista semantico, i formati parzialmente riempiti possono configurarsi come combinazioni dal significato più o meno trasparente. Tuttavia, anche se il livello di composizionalità semantica o di idiomaticità può variare al variare delle strutture combinatorie, gli esempi raccolti per il dizionario *CombiNet* sono caratterizzati spesso da semantica non composizionale:

- (14) Formato sintagmatico/semantico: [*passare per* + Agg_{connotazione negativa}]
 Semantica della combinazione: *essere considerato* X_{AGG}
 Esempi: passare per **fesso**, passare per **imbecille**,
 passare per **matto**, passare per **pazzo**,
 passare per **razzista**, passare per **scemo**
- (15) Formato sintagmatico/semantico: [*prendere a* + Nome_{STRUMENTO(plurale)}]
 Semantica della combinazione: *colpire con* X_{NOME}
 Esempi: prendere a **bastonate**, prendere a **coltellate**, prendere a **martellate**, prendere a **spallate**, prendere a **sassate**

3.1.3. Struttura

I formati parzialmente riempiti possono essere sintagmi più o meno complessi dal punto di vista strutturale e possono avere dei livelli di lessicalizzazione e di fisicità sintattica variabili.

In particolare, i formati parzialmente riempiti possono variare per il numero di costituenti sottoposti a variabilità paradigmatica: i formati possono essere caratterizzati da una (16) o più posizioni variabili (17):

- (16) Formato sintagmatico/semantico: *avere il/la* + Nome + *facile*
 Semantica della combinazione: *essere incline a fare qualcosa connesso con* X_{NOME}
 Esempi: avere il **bicchiere** facile, avere il **grilletto** facile, avere la **lacrima** facile, avere la **pistola** facile, avere la **battuta** facile
- (17) Formato sintagmatico/semantico: [[*di* + Nome_{1[UNITÀ DI TEMPO]}]
 [*in* + Nome_{1[UNITÀ DI TEMPO]}]]
 Semantica della combinazione: *progressione temporale*
 Esempi: di **giorno** in **giorno**, di **minuto** in **minuto**, di **momento** in **momento**, di **anno** in **anno**, di **ora** in **ora**

E' interessante notare che gli stessi formati di parola possono caratterizzarsi per la selezione di *range* paradigmatici diversi. Questo può non influire sulla semantica

della combinazione, come negli esempi (18)–(19), in cui nonostante la variazione lessicale, il formato mantiene la semantica di “dispositivo che funziona per mezzo di X_{NOME} ” (cfr. Piunno 2016a):

- (18) Formato sintagmatico/semantico: [Nome_{DISPOSITIVO} a Nome_{SOSTANZA NATURALE}]
 Semantica della combinazione: *dispositivo che funziona per mezzo di X_{NOME}*
 Esempi: turbina a **gas**, automobile a **benzina**, treno a **vapore**, mulino a **vento**
- (19) Formato sintagmatico/semantico: [NOME_{DISPOSITIVO} a NOME_{CONCRETO}]
 Semantica della combinazione: *dispositivo che funziona per mezzo di X_{NOME}*
 Esempi: barca a **remi**, veicolo a **motore**, pupazzo a **molla**

In altri casi, la variazione lessicale è associata a una variazione semantica. Si confrontino a tal proposito gli esempi seguenti con quelli citati in (11):

- (20) Formato sintagmatico/semantico: [colpo di NOME_{STRUMENTO}]
 Semantica della combinazione: *movimento o azione rapida, svolta attraverso l'uso di $X_{\text{NOME_STRUMENTO}}$*
 Esempi: colpo di **piccone**, colpo di **martello**, colpo di **forbice**, colpo di **pennello**, colpo di **scopa**, colpo di **spazzola**

Mentre le strutture in (11) designano esplosioni, quelle in (20) indicano dei movimenti rapidi. Si può inoltre osservare che alcuni degli esempi in (20) sono oggi giorno lessicalizzati e hanno acquisito una semantica peculiare, spesso altamente idiomatica (es. *colpo di forbice*).

4. Rappresentazione lessicografica dei formati parzialmente riempiti

Data la loro notevole estensione quantitativa e l'interesse teorico del fenomeno, si è scelto di includere nel layout lessicografico del dizionario *CombiNet* anche i formati parzialmente riempiti, a cui è stata dedicata un'apposita sezione dell'entrata lessicografica (Piunno 2016b; Simone e Piunno 2017).

Come nei dizionari tradizionali, ciascuna entrata di *CombiNet* è articolata in diverse accezioni, contenenti una breve glossa con la spiegazione del significato del lemma (fig. 1). Ciascuna accezione contiene informazioni relative alle proprietà combinatorie del lemma. Il layout lessicografico di *CombiNet* è pensato per includere contemporaneamente l'informazione sui diversi tipi sintagmatici e sui numerosi formati di parola identificati, ed è articolato in quattro colonne (fig. 2) (Simone – Piunno 2017):

The screenshot shows the CombiNet interface for the lemma 'colpo'. At the top, there is a navigation bar with 'Home / Lemma / colpo'. Below this, the lemma 'colpo' is identified as a masculine singular noun ('sostantivo m. s.'). There are three buttons: 'Indietro', 'Mostra tutte', and 'Crea nuova accezione'. The main content area lists five numbered entries:

1. Urto, percossa; *estens.* rumore prodotto dall'urto.
2. Manifestazione, azione o movimento rapido e improvviso.
- 3.a Evento improvviso, solitamente dannoso. Spavento.
- 3.b Attacco improvviso, stato morboso.
- 3.c Azione rapida, perlopiù illegale.

Figura 1. Entrata lessicografica di *colpo* in *CombiNet*.

This screenshot shows a detailed view of the lemma 'colpo' entry. It features the same navigation and buttons as Figure 1. The first entry is expanded to show a table with four columns: 'Categoria', 'Struttura', 'Dati primari', and 'Dati secondari'. Below the table, it indicates 'Elementi totali 1.'.

Categoria	Struttura	Dati primari	Dati secondari
Elementi totali 1.			

Figura 2. Articolazione del layout lessicografico in quattro colonne.

Le prime due colonne contengono le informazioni sintattiche relative alle unità combinatorie. La prima colonna include informazioni sulla categoria di uscita (la funzione che una combinazione di parole assume in contesto sintagmatico, definita in *CombiNet* con l'etichetta "Categoria"), la seconda il formato sintagmatico della combinazione, espresso come strutture sintagmatiche (registrata con l'etichetta "Struttura").

La terza e la quarta colonna registrano gli esempi combinatori estratti dal corpus e ne specificano le informazioni semantiche. La terza colonna ("Dati Primari") è dedicata alla registrazione degli esempi di formati completamente riempiti, mentre la quarta ("Dati Secondari") registra quelli parzialmente riempiti. Per ciascun formato parzialmente riempito *CombiNet* registra il formato di parola (struttura sintagmatico-semantic), la semantica e gli esempi combinatori estratti dal corpus (fig. 3) (Simone – Piunno 2017):

colpo sostantivo m. s.

◀ Indietro Mostra tutte ▶ Crea nuova associazione

1. Urto, percossa; *estens.* rumore prodotto dall'urto.

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

Figura 3. Esempi combinatori per l'entrata *colpo*.

5. Riflessioni conclusive

La compilazione delle voci lessicografiche per *CombiNet* ha messo in luce nuove proprietà e nuovi possibili parametri classificatori delle unità combinatorie. Questo ha generato alcune riflessioni in merito al concetto di *combinazioni di parole*. In particolare è emerso che:

- i. le combinazioni di parole costituiscono un insieme variegato e intensamente popolato, che va analizzato secondo parametri diversi rispetto a quelli tradizionalmente usati in letteratura;
- ii. le combinazioni di parole vanno distinte in *completamente* e *parzialmente riempite*, sulla base del grado di variabilità lessicale e della produttività dei formati di parola che le caratterizzano;
- iii. le combinazioni parzialmente riempite hanno lo stesso statuto delle combinazioni completamente riempite, sebbene mostrino proprietà diverse;
- iv. i profili combinatori previsti in letteratura devono necessariamente essere affinati attraverso l'esame dei dati, anche tenendo conto dei diversi gradi di produttività dei vari formati di parola.

A questo può aggiungersi che, dal punto di vista lessicografico, un'opera dedicata alla rappresentazione della combinatoria lessicale non può prescindere dal considerare i molteplici fenomeni di regolarità che caratterizzano il lessico di una lingua. I diversi tipi combinatori (inclusi i formati parzialmente riempiti) meritano pertanto l'inclusione in un'opera lessicografica di questo tipo. E' inoltre importante notare che l'identificazione di nuovi fenomeni di regolarità nel lessico può avere delle importanti ripercussioni sul piano computazionale e sul trattamento automatico dei dati linguistici.

Bibliografia

- Baroni M., Bernardini S., Comastri F., Piccioni L., Volpi A., Aston G., Mazzoleni M. 2004. Introducing the *La Repubblica* corpus: a large, annotated, TEI(XML) – compliant corpus of newspaper Italian, [in:] Proceedings of the Fourth International Conference on Language Resources and Evaluation, a c. di M.T. Lino, M.F. Xavier, F. Ferreira, R. Costa, R. Silva, Paris: 1771–1774.

- Bybee J. 2010. *Language, Usage and Cognition*, Cambridge.
- Bybee J., Thompson S. 1997. "Three Frequency Effects in Syntax". *Berkeley Linguistic Society* 23: 65–85.
- Castagnoli S., Lebani G.E., Lenci A., Masini F., Nissim M., Piuino V. 2015. Towards a corpus-based online dictionary of Italian Word Combinations Automatic Knowledge Acquisition for Lexicography, COST ENeL WG3 meeting, Herstmonceux Castle, 13 August 2015. www.elxicography.eu/wp-content/uploads/2015/10/.
- Castagnoli S., Lebani G.E., Lenci A., Masini F., Nissim M., Passaro L.C. 2016. POS-patterns or Syntax? Comparing methods for extracting Word Combinations, [in:] *Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives (Full papers) – Fraseología computacional y basada en corpus: perspectivas monolingües y multilingües (Trabajos completos)*, a c. di G. Corpas Pastor, Geneva: 101–114.
- Fillmore C.J., Kay, P., O'Connor, M.C. 1988. "Regularity and idiomaticity in grammatical constructions: the case of let alone". *Language* 64: 501–38.
- Gross G. 1996. *Les expressions figées en français*, Paris.
- Lamiroy B. 2003. "Les notions linguistiques de figement et de contrainte". *Linguisticae Investigationes* 26 (1): 1–14.
- Lenci A., Lapesa G., Bonansinga G. 2012. LexIt: A Computational Resource on Italian Argument Structure, [in:] *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, ELRA: 3712–3718.
- Masini F. 2013. *Parole sintagmatiche in italiano*, Cesena, Caissa Italia.
- Mejri S. 1997. *Le figement lexical. Descriptions linguistiques et structuration sémantique*, Tunisie.
- Passaro L.C., Lenci A. 2016. Extracting Terms with EXTra, Paper presented at EUROPHRAS 2015 – Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives. Malaga, 29 June–1 July 2015.
- Piuino V. 2013. *Modificatori sintagmatici con funzione aggettivale e avverbale*, Tesi di dottorato, Roma.
- Piuino V. 2015. "Sintagmi Preposizionali come Costruzioni Aggettivali", *Studi e Saggi di Linguistica* 53 (1): 65–98.
- Piuino V. 2016a. "Multiword Modifiers in Romance languages. Semantic formats and syntactic templates". *Yearbook of Phraseology*, Berlin 7: 3–37.
- Piuino V. 2016b. CombiNet. A Corpus-based Online Database of Italian Word Combinations, GLOBALEX 2016: Lexicographic Resources for Human Language Technology, LREC2016, Portorož, 24 maggio 2016: 45–51.
- Sechehaye A. 1950 [1926]. *Essai sur la structure logique de la phrase*, Collection linguistique 20, Paris.
- Simone R. 2006. Nominales sintagmáticos y no-sintagmáticos, [in:] *Estructuras léxicas y estructuras del léxico*, red. E. De Miguel *et al.*, Frankfurt am Main, Peter Lang: 225–246.
- Simone R. 2007. Constructions and categories in verbal and signed languages, [in:] *Verbal and Signed Languages. Comparing Structures, Constructs, and methodologies*, red. Pietrandrea P. *et al.*, Berlino–New York: Mouton-De Gruyter: 198–252.
- Simone R., Piuino V. 2017. "Combinazioni che costituiscono lemma. Fenomeni, rappresentazione lessicografica e aspetti lessicologici", *Studi e saggi linguistici* 55 (2): 13–44.
- Tutin A., Grossmann F. 2002. "Collocations régulières et irrégulières: esquisse de typologie du phénomène collocatif", *Revue française de Linguistique appliquée* 7 (1): 7–25.

Voghera M. 1994. "Lessemi complessi: percorsi di lessicalizzazione a confronto", *Lingua e Stile* 29 (2): 185–214.

Voghera M. (2004. Polirematiche, [in:] *La formazione delle parole in italiano*, M. Grossmann, F. Rainer, Tübingen, Max Niemeyer Verlag: 56–69.

Combinazioni di parole parzialmente riempite. Formati e rappresentazione lessicografica

Questo contributo intende analizzare le proprietà di alcune combinazioni di parole della lingua italiana, rappresentate da strutture riempite solo parzialmente da unità lessicali. Attraverso l'analisi di dati estratti da corpora, il contributo si propone di a) identificare e analizzare alcuni *formati parzialmente riempiti* dell'italiano emersi durante la stesura del dizionario combinatorio *CombiNet*, b) classificare tali unità sulla base di specifiche proprietà sintattiche o semantiche, c) mostrare una possibile rappresentazione lessicografica di tali formati di parola.

Parole chiave: combinazioni di parole, combinazioni parzialmente riempite, formato di parola, *CombiNet*, dizionario combinatorio

Partially filled word combinations. Formats and lexicographic representation

The work here proposed is intended as a description of word combinations which are *partially filled* by lexical units. The analysis of real corpus data, the investigation aims at a) identifying and analysing some Italian partially filled word combination emerged during the development of *CombiNet* combinatory dictionary, b) classifying *partially filled units* on the basis of their syntactic and semantic properties, c) showing a specific lexicographical representation of such word formats.

Keywords: word combinations, partially filled word combinations, word format, *CombiNet*, combinatory dictionary of Italian

Kombinacje słów częściowo wypełnione. Formacje oraz reprezentacja leksykograficzna

Niniejszy artykuł ma na celu przeprowadzenie analizy właściwości niektórych kombinacji słów w języku włoskim przedstawionych za pomocą struktur, które zostały „wypełnione” jedynie w sposób częściowy przez jednostki leksykalne. Artykuł, poprzez analizę danych uzyskanych z korpusów, proponuje: a) identyfikację oraz analizę niektórych formacji częściowo wypełnionych w języku włoskim, które zostały zidentyfikowane podczas sporządzania słownika kombinatorycznego *CombiNet*; b) klasyfikację danych jednostek na podstawie określonych właściwości syntaktycznych bądź semantycznych; c) przedstawienie możliwej reprezentacji leksykograficznej danych formacji słów.

Słowa kluczowe: kombinacje słów, kombinacje częściowo „wypełnione”, formacje słów, *CombiNet*, słownik kombinatoryczny

Valentina Piunno – è assegnista di ricerca presso l'Università degli Studi Roma Tre. Ha conseguito il dottorato in Linguistica Sincronica, diacronica e applicata nel 2013 presso l'Università degli Studi Roma Tre, discutendo una tesi sui "Modificatori sintagmatici con funzione aggettivale e avverbiale". Ha collaborato al progetto "CombiNet", occupandosi della creazione di un *tool* lessicografico per la rappresentazione delle combinazioni di parole. I suoi interessi di ricerca si rivolgono principalmente allo studio del lessico (italiano, spagnolo e francese) e alla combinatoria lessicale, alla lessicografia combinatoria, alla semantica lessicale, alla linguistica dei corpora.