

# LE COMBINAZIONI DI PAROLE PARZIALMENTE RIEMPITE IN ALCUNE LINGUE ROMANZE. SCHEMATISMO E PREDICIBILITÀ SEMANTICA

Valentina Piunno

Dipartimento di Studi linguistico-letterari, storico-filosofici e giuridici (DISTU),  
Università degli Studi della Tuscia, Via San Carlo 32, 01100 Viterbo, Italia  
valentina.piunno@unitus.it

## Partially filled constructions in some Romance languages. Schematicity and semantic predictability

**Abstract:** This investigation proposes a *corpus-driven* description of *partially filled constructions* (PFCs), a subtype of phraseological units characterized by a low degree of lexical specification and a high level of semantic predictability. PFCs contain both fixed and empty slots, showing variable levels of productivity. As *constructions*, they are pairings of form and function (Goldberg), since their semantic/syntactic configuration can convey particular sets of meanings. Through the analysis of authentic data extracted from *corpora* of three different Romance languages (Italian, French, and Spanish), this contribution aims at i) identifying and analysing a set of partially filled units, ii) showing a formal representation of such syntactic-semantic patterns, iii) classifying them into different types on the basis of functional, syntactic, and semantic criteria, as well as on the degree of idiomaticity and lexical variation, and iv) driving a cross-linguistic comparison among the Romance languages. The same PFC may be characterized by different levels of schematicity and productivity in the various languages taken into account. It will be shown that PFCs may be associated with different syntactic structures, either at the phrase-level or at the sentence-level. Even though they are characterized by empty slots and by a lower degree of lexical specification, they still represent a regular area of the lexicon. New and independent lexemes can develop from specific constructions and become conventionalized through repetition.

**Keywords:** partially filled constructions; lexical variability; productivity; empty slots; Romance languages

**Riassunto:** Questo contributo propone una descrizione *corpus-driven* di un tipo di strutture combinatorie, accomunate da specifiche proprietà morfosintattiche e semantiche: le *combinazioni parzialmente riempite* (CPR). Le CPR sono caratterizzate da un basso grado di specificità lessicale, a cui corrisponde un'alta predicibilità semantica; si configurano attraverso un

formato superficiale caratterizzato da posizioni vuote, con una caratterizzazione semantica stabile e un livello di produttività variabile. Al pari delle *costruzioni*, le CPR rappresentano un'associazione di forma e significato (Goldberg), e sono in grado di veicolare specifici significati attraverso la configurazione sintattico/semantica. Le CPR possono essere classificate sulla base di diversi parametri, tra cui le proprietà funzionali, il grado di idiomatichità, il livello di specificità lessicale e la produttività della struttura. Attraverso l'analisi di dati estratti da *corpora*, il contributo si propone di identificare e analizzare i formati parzialmente riempiti in alcune lingue di ambito romanzo (italiano, spagnolo e francese), e di mostrarne una loro possibile rappresentazione formale e una classificazione in tipi, tenendo conto delle eventuali differenze interlinguistiche. Si noterà come le stesse CPR nelle diverse lingue possono essere caratterizzate da tratti di schematicità e livelli di produttività variabili. Infine, si dimostrerà che le CPR rappresentano un fenomeno di regolarità del lessico, nonostante siano caratterizzate da un basso grado di specificità lessicale. Nuovi elementi lessicali possono originarsi a partire da una specifica costruzione, e convenzionalizzarsi attraverso la ripetizione.

**Parole-chiave:** combinazioni parzialmente riempite; variabilità lessicale; produttività; posizioni vuote; lingue romanze

## 1. Introduzione: premessa teorica e terminologica<sup>1</sup>

Numerosi recenti studi di ordine lessicologico e lessicografico hanno mostrato che il lessico delle lingue non è costituito da elementi utilizzati in isolamento, ma da unità attorno alle quali vengono create relazioni di diversa natura, di diversa intensità e forza coesiva (Simone 2007). Alcune combinazioni di parole possono dissolversi al momento dell'enunciazione, mentre altre possono lessicalizzarsi e, col tempo, dar vita a nuova designazione. Nell'ambito della combinatoria lessicale troviamo un'ampia gamma di fenomeni combinatori, che mostrano diversi gradi di cristallizzazione sintattica, idiomatichità e lessicalizzazione. I diversi tipi combinatori possono essere posti lungo un *continuum* che va dalla sintassi al lessico, e che include pertanto i) fenomeni ascrivibili alla sintassi ma non rintracciabili nel lessico (polo sinistro del *continuum*) e ii) fenomeni che hanno origine in sintassi, ma che col tempo si fissano ed entrano a far parte del lessico di una lingua (polo destro del *continuum*) (cfr. Simone 2007: 212). Al centro del *continuum* si trovano combinazioni di parole dotate di variabile intensità coesiva (Simone 2007) e fissità lessicale. Nella figura che segue sono rappresentate, a titolo esemplificativo, le diverse posizioni del *continuum* identificate in Simone (2007) (per es. combinazioni preferenziali, collocazioni, costruzioni propriamente dette, ecc.).

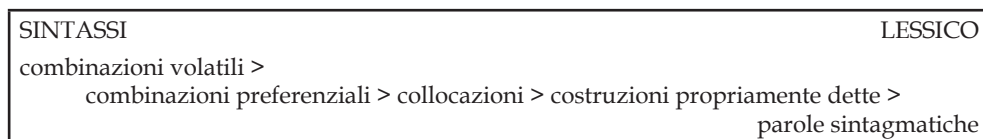


Figura 1. Derivazione diacronica delle costruzioni nel continuum sintassi-lessico (Simone 2007: 212, traduzione mia)

<sup>1</sup> Questo contributo costituisce un'estensione e una rielaborazione di quanto riportato in Piunno (2016a, 2018).

Lungo il *continuum* sintassi-lessico è possibile porre, pertanto, combinazioni di parole caratterizzate da un diverso grado di specificità lessicale e rigidità sintattica. In considerazione di tali parametri, è possibile disporre ai poli opposti del *continuum* le combinazioni di parole *completamente riempite* e quelle che lo sono solo *parzialmente*:

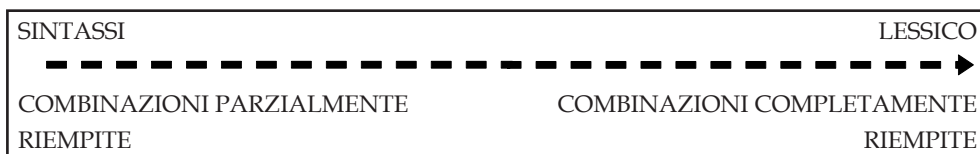


Figura 2. Combinazioni “parzialmente” e “completamente” riempite nel continuum sintassi – lessico

Nel primo caso la sequenza combinatoria presenta un alto grado di lessicalizzazione, è interamente specificata dal punto di vista lessicale (vale a dire, le proprietà lessicali della combinazione sono stabili e invariabili), mostra un significato non prevedibile a partire da quello dei singoli costituenti (spesso idiomatico), e non ammette variazione sintattica. Tipici esempi di questa tipologia di combinazioni sono le costruzioni a verbo supporto (per es. *dare una mano, fare benzina*), le parole sintagmatiche<sup>2</sup> (per es. *luna di miele, tirare su, mandare in onda, a senso unico, per mezzo di, dal momento che*), le costruzioni binomiali (per es. *mordi e fuggi*), le espressioni idiomatiche (per es. *non vedere l'ora*), le interiezioni (per es. *grazie al cielo!*) e i proverbi (per es. *occhio non vede cuore non duole*).

Dall'altro lato, le combinazioni *parzialmente riempite*<sup>3</sup>, oggetto della presente analisi, sono caratterizzate dalla presenza di un pattern variabile, vale a dire soggetto in misura parziale alla variazione lessicale. In particolare, tali sequenze si compongono di una parte stabile (fissa e immutabile), e di una parte soggetta a variazione. Si osservi, a titolo esemplificativo, la sequenza che segue:

(1) *prendere a* bastonate/ sassate/ martellate/ testate / sprangate / pallonate/ picconate

La sequenza in (1) è caratterizzata dalla presenza i) di una porzione invariabile (contrassegnata in corsivo, i.e. *prendere a*) e di ii) una porzione soggetta a variazione lessicale e sottoposta a specifiche restrizioni (segnalata in tondo). In particolare, l'elemento lessicale richiesto a completamento del sintagma verbale risponde a specifici criteri semantici e morfosintattici:

- i) deve necessariamente essere un nome, e più in particolare, un nome che indica un evento – un *nome di una volta* (cfr. Simone 2003) derivato dalla combinazione di un nome di strumento con il suffisso *-ata -*,
- ii) deve essere flesso al plurale, per indicare la reiterabilità dell'evento.

<sup>2</sup> Il termine include i nomi sintagmatici, i verbi complessi e i verbi sintagmatici, gli aggettivi sintagmatici, gli avverbi sintagmatici, le preposizioni sintagmatiche e le congiunzioni sintagmatiche. Per approfondimenti si rimanda, in particolare, a Voghera (1994, 2004) e Masini (2013).

<sup>3</sup> Con il termine “costruzioni parzialmente riempite” si traduce il concetto denominato *partially lexically-filled constructions* in Goldberg (2006) e nella letteratura americana contemporanea di stampo costruzionista (cfr. § 2).

In questo caso specifico, la costruzione [*prendere a* + N<sub><strumento>-ate</sub>] acquisisce il significato di ‘colpire ripetutamente con X<sub>N1</sub>’.<sup>4</sup>

Le CPR sono relativamente più *flessibili* delle combinazioni completamente riempite. Le posizioni vuote possono variare nel numero e nel tipo, a seconda della struttura. Pur essendo flessibili (o variabili), dal punto di vista lessicale mantengono comunque una certa regolarità: ammettono infatti la saturazione con materiale della stessa natura morfosintattica o semantica. Dal punto di vista semantico, possono avere un significato prevedibile (anche quando idiomatico) a partire dall’intera costruzione. Infine, sulla base del numero di esemplari che da esse possono essere ricavati, le CPR possono mostrare diversi livelli di produttività (Bybee 2010: 94).

Pur non rappresentando delle unità multiparola prototipiche<sup>5</sup>, poiché dotate di un basso grado di coesione e di fissità sintattica, e un minore livello di lessicalizzazione rispetto alle parole sintagmatiche, i formati parzialmente riempiti sembrano comunque rappresentare un fenomeno di regolarità nel lessico. In particolare, riprendendo la terminologia utilizzata in Simone (2007),<sup>6</sup> le CPR costituiscono un tipo particolare di *formato di parola* (Simone 2017), rappresentato da un *pattern* sintattico-semantico in grado di veicolare specifici valori semantici o pragmatici. Le CPR sono strutture schematiche, astratte, caratterizzate dalla presenza di alcune posizioni non specificate a livello lessicale. I formati parzialmente riempiti rientrano nella nozione di *costruzione* generata in ambito americano verso la fine del secolo scorso<sup>7</sup> e adottata in ambito europeo con sfumature variabili<sup>8</sup> (cfr. § 2). In particolare, al pari delle *costruzioni*<sup>9</sup>, le strutture parzialmente riempite sono «learned pairings of form and (semantic or discourse) function» (Goldberg 2003: 220), in grado di rappresentare in un’unica struttura l’informazione sintattica e semantica ad essa associate (Simone 2007). Così come le costruzioni *tout court*, esse sono strutture più ampie della parola, che però sono «like words in that they have to be learned separately as individual whole facts» (Fillmore *et al.* 1988: 504). Inoltre, pur essendo solo parzialmente specificate, possono comunque veicolare uno specifico significato. Le CPR possono persino presentare un significato idiomatico, «in the sense that a large construction may specify a semantics (and/or pragmatics) that is distinct from what might be calculated from the associated semantics of its

<sup>4</sup> Vale la pena notare che la stessa semantica può essere veicolata da strutture simili. Si prenda, ad esempio, la struttura [*prendere a* + nomi di evento-plurale] (es. *prendere a calci, prendere a colpi di bastone*, ecc.). In questo caso si ha, tuttavia, un basso livello di produttività, poiché il completamento della sequenza ammette un limitato numero di elementi lessicali a saturazione delle posizioni vuote.

<sup>5</sup> Per questo concetto si rimanda a Lamiroy (2003), Simone (2007), Svensson (2004), Masini (2013), Piunno (2018) e ai riferimenti ivi citati.

<sup>6</sup> Cfr. *Categories and Constructions Grammar* (Simone 2007, 2017).

<sup>7</sup> Cfr. *Construction Grammar* (tra gli altri, in particolare, Fillmore *et al.* 1988; Goldberg 2003, 2006; Michaelis, in stampa a, b).

<sup>8</sup> Cfr. ad esempio la nozione di “costruzione” in Simone (2007, 2017) e quella adottata in ambito fraseologico da Dobrovolskij (2011), Steyer (2018) e Mellado Blanco - López Meirama (2018).

<sup>9</sup> Con *costruzione* si fa riferimento, in particolare, al concetto introdotto nell’ambito degli studi di *construction grammar* a partire da Fillmore *et al.* (1988).

components» (Fillmore *et al.* 1988: 501). Inoltre, costituiscono delle configurazioni parzialmente schematiche: il pattern sintattico/semantico che le caratterizza può mostrare diversi livelli di produttività, in termini di nuove forme da esso derivabili (Bybee 2006, 2010).

In quest'ottica, le CPR costituiscono dunque un tipo specifico di *formato di parola* (Simone 2007, 2017), ovvero una struttura linguistica sintatticamente complessa in grado di rappresentare al contempo un'informazione di natura sintattica e una di natura semantica. In particolare, tra i diversi formati di parola di cui la lingua dispone, i formati parzialmente riempiti possono rientrare nel cosiddetto *formato multiparola*. Quest'ultimo caratterizza tutte le sequenze lessicali con struttura sintagmatica (cioè composte da più parole formanti un'unità, per es. *luna di miele, mettere bocca, in bianco e nero*), i cui costituenti sono tenuti insieme da una *forza coesiva* di vario grado (Simone 2007).

## 2. Breve stato dell'arte

Sin dagli inizi del Novecento, molti linguisti hanno dedicato gran parte dei loro studi all'analisi delle combinazioni di parole e alla loro capacità di costituire una nuova designazione. Se, da un lato, i diversi studi hanno prediletto l'analisi delle regolarità attribuibili alle combinazioni sintagmatiche fisse (alle quali è stata dedicata una vasta molteplicità di studi), dall'altro lato non sono passate inosservate le combinazioni di parole dal carattere più astratto, ma generate da schemi altrettanto regolari. In particolare, molti studiosi hanno postulato l'esistenza di diversi tipi combinatori, a cui corrispondono spesso diversi gradi di lessicalizzazione.

Nell'ambito della linguistica svizzera di stampo saussuriano, ad esempio, si trovano riferimenti alla distinzione tra unità fraseologiche *dalla maggiore flessibilità sintattica* e quelle *totalmente invariabili*. Analisi di questo tipo si trovano nel *Traité de stylistique française* di Bally (1951[1909]) in cui i *groupements de mots* sono distinti in *serie fraseologiche* e *unità fraseologiche*, sulla base dei diversi gradi di cristallizzazione sintattica. Le prime, al pari delle CPR, sono dotate di una maggiore flessibilità sintattica, che ne consente la sostituzione delle parti con elementi della stessa classe, mentre le seconde non ammettono scambi o sostituzioni di costituenti. Altri studiosi, pur prediligendo la caratterizzazione sintattica di tali sequenze, chiamano in causa le loro proprietà lessicali. Secondo Frei (1969[1929]),<sup>10</sup> ad esempio, alcune combinazioni di parole ammettono la sostituzione parziale con un loro sinonimo o un loro antonimo.<sup>11</sup> Su queste basi, in ambito funzionalista, Pottier (1974[1967]) introduce il concetto di *degré de lexicalisation*, distinguendo tra *lexies figées* e *lexies variables*. I due tipi combinatori, secondo Pottier, sono caratterizzati da un diverso grado di variabilità lessicale. Mentre i primi sono sequenze totalmente invariabili

<sup>10</sup> Va tuttavia sottolineato che, nel suo *La grammaire des fautes*, Frei (1969[1929]), si limita a individuare alcuni segnali utili all'identificazione dei diversi casi di lessicalizzazione, non stabilendo dei veri e propri criteri di classificazione.

<sup>11</sup> Tale operazione non è di norma ammessa nelle sequenze fisse (per es. *bon marché* non può essere sostituito con *mauvais marché*).

in via di lessicalizzazione (per es. *point d'honneur, à vol d'oiseau*), le seconde sono sequenze caratterizzate da una zona stabile e da una zona instabile; quest'ultima è la porzione che ammette, anche se in maniera limitata, la variazione lessicale (per es. *le feu {vert/rouge/orange/clignotant}*, *le {premier/dernier/nouveau} né*) (Pottier 1985: 267).

Più recentemente, l'ambito delle strutture che sono solo parzialmente specificate da materiale lessicale è stato oggetto di interesse degli studi di *Construction Grammar*, fortemente incentrati sul tema dell'interazione tra sintassi e lessico, e sul *continuum* che lega i due livelli di analisi.<sup>12</sup> In particolare, ciò che accomuna molti lavori di stampo costruzionista è «the idea that constructions are learned pairings of form with semantic or discourse function, including morphemes (e.g. *un, -er*), words (e.g. *dog, run*), filled idioms (e.g. *kick the bucket*), partially filled idioms (e.g. *a pain in the X*), partially lexically filled phrasal patterns (e.g. *passive constructions*), and fully general phrasal patterns (e.g. *subject-predicate agreement constructions*)» (Boas 2010b: 3). Questo vuol dire che le diverse costruzioni possono differenziarsi sulla base del grado di schematicità e/o di specificità lessicale. Inoltre, come specificato in Goldberg (2006: 5), «[a]ny linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist». Già a partire dagli albori del costruzionismo, infatti, le combinazioni di parole sono state classificate infatti in termini di flessibilità sintattica e lessicale, in diverse tipologie. Nella classificazione proposta in Fillmore *et al.* (1988), emergono in particolare le classi dei *substantive* e *formal idioms*; i primi sono completamente riempiti e fissi in termini lessicali (per es. *kick the bucket, spill the beans, blow one's nose*), i secondi costituiscono dei *frames* sintattici che possono veicolare una semantica lessicale o pragmatica specifica (per es. la costruzione inglese *let alone*, cfr. § 4.1). Nell'ambito del quadro teorico costruzionista, numerosi altri studi hanno riconosciuto l'esistenza di diversi sottotipi di costruzioni o sono stati dedicati a questo tipo specifico di combinazioni di parole. Tra questi, citiamo la classificazione presentata in Goldberg (2006), nella quale le diverse costruzioni includono pattern variabilmente specificati a livello lessicale; in particolare, tra le diverse strutture identificate, quelle che più si avvicinano al tema oggetto della presente analisi sono le *partially lexically-filled constructions*, rappresentate da espressioni idiomatiche flessibili sul piano lessicale. Si tratta di strutture che presentano un alto grado di idiomatizzazione semantica e, di norma, una posizione sintattica variabile: es. *send <someone> to the cleaners* ("ridurre <qualcuno> sul lastrico"). In Michaelis (in stampa a, b) e in Kay - Michaelis (2012) le combinazioni di parole parzialmente riempite (classe in cui possono rientrare le *snow clones*)<sup>13</sup> rappresentano una categoria intermedia nel

<sup>12</sup> Cfr. Boas (2010a) per una discussione sul *continuum* tra i due livelli di analisi secondo il quadro teorico della *Construction Grammar*.

<sup>13</sup> Uno *snow clone* è una formula fissa, di norma derivata da una citazione, che viene modificata per scopi pragmatici precisi (per es. giochi di parole, marketing, ecc.) come ad esempio (Michaelis in stampa, a: 2):

(1) "I  $X_{\text{VERBO}}$  therefore I am"  $X_{\text{VERBO}}$  dunque sono'

L'espressione in (1), derivata dall'espressione cartesiana «Cogito ergo sum», presenta una posizione

*continuum* che vede ai poli opposti le combinazioni lessicali fisse e le configurazioni sintattiche più astratte (Kay – Michaelis 2012). Secondo questa cornice teorica, le combinazioni di parole parzialmente riempite costituiscono delle vere e proprie espressioni idiomatiche *flessibili* «in the sense that the patterns on which they are based can be used to produce novel instances» (Michaelis in stampa, a: 2). In questo caso le costruzioni sono delle strutture sintagmatiche preconfezionate (*phrasal templates*), caratterizzate dalla presenza di una porzione fissa e alcune posizioni aperte alla variazione lessicale.

In ambito europeo, molti studi di fraseologia fortemente legati ai quadri teorici di stampo costruzionista hanno recentemente individuato diversi fenomeni caratterizzati da un variabile grado di specificità lessicale e idiomaticità costruzionale<sup>14</sup>. Tra le più recenti, citiamo le analisi quantitative basate su corpus di Ziem (2008, 2018) e Ziem – Lasch (2013), volte a descrivere il rapporto tra i diversi gradi di schematicità e la semantica del *frame*. In particolare, in Ziem (2018) sono elencati i diversi parametri utili all'analisi di strutture costruzionali, tra cui: i) la semantica complessiva (spesso idiomatica), ii) la presenza di restrizioni specifiche, iii) la presenza di posizioni vuote e la frequenza di *fillers* tipici, iv) la produttività e v) le relazioni con altre costruzioni. Le strutture fraseologiche con uno schema sintagmatico fisso e posizioni sintattiche vuote (i cui costituenti possono essere specificati lessicalmente) possono essere caratterizzate da gradi diversi di produttività: ad un alto numero di elementi lessicali ammessi a saturazione delle posizioni vuote corrisponde una elevata produttività (Ziem 2018: 10). Nello stesso ambito, molti studi tendono a sottolineare lo stretto rapporto tra unità fraseologiche e costruzioni, e a concentrarsi sulla comune presenza di schemi sintattici che nascondono specifiche peculiarità semantiche. Secondo questi approcci tali costruzioni rappresentano dei pattern multiparola (*multi-word patterns*, cfr. Steyer 2018), ai confini tra le strutture fraseologiche e le costruzioni *tout court*. In particolare, in Steyer (2018) i cosiddetti *pattern multiparola* sono strutture con posizioni vuote, che condividono alcune proprietà semantiche (possono essere saturati da materiale lessicale semanticamente omogeneo) e pragmatiche. Alla stregua delle strutture fraseologiche, i pattern multiparola rappresentano delle unità lessicali dotate di significato e funzione specifici. Un simile quadro teorico può ritrovarsi negli studi di Dobrovól'skij (2011), secondo il quale le costruzioni possono presentarsi come unità schematiche complesse (le cosiddette *Phrasem-Konstruktionen*), caratterizzate da una struttura sintattica con porzioni lessicalmente riempite e posizioni vuote. Queste ultime sono di norma soggette a specifiche restrizioni semantiche, strettamente connesse al significato costruzionale. Di forte ispirazione costruzionista sono anche le analisi delle *construcciones fraseológicas* operate in Mellado Blanco – López Meirama (2017, 2018) e in Mellado Blanco (in stampa). In particolare, in queste ultime si sottolinea la natura flessibile delle strutture parzialmente riempite, sia per il grado di schematicità sia dal punto di vista della produttività. Gli studi mostrano il

vuota, soggetta a possibile variazione.

<sup>14</sup> Per questo concetto si rimanda a Piunno (2015).

rapporto direttamente proporzionale tra produttività e flessibilità lessico/sintattica. In particolare, ad un maggior livello di produttività corrisponde in genere un numero minore di restrizioni lessicali, morfologiche e sintattiche applicabili alla costruzione. Secondo questo approccio la classe delle *construcciones fraseológicas* non possiede confini definiti: grazie all'uso, alcuni schemi sintagmatici attivano specifiche attualizzazioni lessicali, che possono lessicalizzare dando vita a locuzioni fisse (Mellado Blanco – López Meirama 2018: 577). In questo ambito, gli studi più recenti tendono a rimarcare la stretta correlazione tra la frequenza d'uso di alcune strutture e i) il radicamento di alcuni schemi sintagmatici da un lato (cfr. Bybee 2006; 2010), e ii) la cristallizzazione di sequenze specifiche dall'altro (Mellado Blanco – López Meirama 2018).<sup>15</sup>

### 3. Estrazione dei dati e metodologia di analisi

Il lavoro è incentrato sull'analisi sincronica delle CPR in italiano, spagnolo e in francese. Il campione di dati utilizzato per lo scopo della presente analisi è stato estratto attraverso *corpora* (o strumenti di analisi) specifici, disponibili per ciascuna lingua considerata. L'estrazione dei dati nelle diverse lingue è stata effettuata seguendo metodi diversi, sulla base del campione di lingua considerato.

I formati di parola parzialmente riempiti dell'italiano provengono dal materiale raccolto in Piunno (2018) e da un campione selezionato nell'ambito della creazione delle voci lessicografiche per il dizionario *CombiNet*<sup>16</sup>. Le CPR sono state inizialmente individuate attraverso l'estrazione dei pattern combinatori più frequenti della lingua italiana, in termini di sequenze contigue di classi di parole preselezionate (ovvero *PoS-grams* di massimo 4 posizioni). Le configurazioni sintagmatiche predefinite (sequenze specifiche di parti del discorso identificate dai più recenti studi<sup>17</sup> sulle combinazioni di parole in italiano) sono state selezionate in base ad indici statistici che misurano il livello di combinatorietà<sup>18</sup> e i risultati sono stati ordinati secondo una specifica misura statistica di associazione (la *log likelihood*). Di seguito un esempio dei pattern estratti per il nome *mano*<sup>19</sup>:

<sup>15</sup> Alcune strutture mostrano caratteristiche morfosintattiche e proprietà semantiche simili. Grazie all'uso, alcune di esse si fissano sintatticamente, fino a divenire vera e propria lessicalizzazione. Queste costituiscono gli "insiemi sintetici" cui fa riferimento Sechehaye (1950[1926]: 97).

<sup>16</sup> Il dizionario *CombiNet* è un database lessicografico che raccoglie alcune combinazioni di parole della lingua italiana. I dati in esso inclusi sono stati estratti da due *corpora* di italiano – il *corpus* giornalistico *La Repubblica* (Baroni *et al.* 2004) e il *corpus* di italiano del web *Paisà* (*Piattaforma per l'Apprendimento dell'Italiano su Corpora Annotati*, Lyding *et al.* 2014) – per mezzo del software *EXTra* (cfr. Lenci *et al.* 2017; Passaro – Lenci 2016; Castagnoli *et al.* 2015). Per approfondimenti si rimanda a Simone – Piunno (2017), Piunno (2016c).

<sup>17</sup> Cfr., tra gli altri, Voghera (1994; 2004), Simone (2006, 2007, 2008), Simone – Masini (2007; 2014), Masini (2013), Piunno (2015; 2016b; 2018).

<sup>18</sup> Cfr. per approfondimenti Castagnoli *et al.* (2015, 2016), Passaro – Lenci (2016), Lenci *et al.* (2017), Simone – Piunno (2017)..

<sup>19</sup> Cfr. Passaro – Lenci (2016).



LL	Fq	parola 1	PoS_1	parola 2	PoS_2	parola 3	PoS_3	parola 4	PoS_4
218.693.481.805	1816	man	s	mano	s				
16.958.368.682	1499	portata	s	di	e	mano	s		
165.195.988.026	1478	stretta	s	di	e	mano	s		
119.794.932.506	764	mano	s	in	e	mano	s		
119.794.932.506	942	mano	s	a	e	mano	s		
119.794.932.506	286	mano	s	nella	ea	mano	s		
464.978.803.663	638	bomba	s	a	e	mano	s		
463.197.815.616	982	colpo	s	di	e	mano	s		
400.401.311.881	106	dita	s	della	ea	mano	s		
400.401.311.881	291	dita	s	di	e	una	ri	mano	s
305.306.992.211	513	chiave	s	in	e	mano	s		
183.566.155.637	105	palmo	s	della	ea	mano	s		
114.119.861.115	140	freno	s	a	e	mano	s		
925.607.882.865	48	pistola	s	alla	ea	mano	s		

Tabella 1. Pattern estratti per la parola *mano* attraverso il software EXTra (Passaro – Lenci 2016)

I diversi esemplari relativi a ciascuna CPR sono stati successivamente estratti attraverso il software *SketchEngine* dal corpus di italiano scritto *itTenTen16*<sup>20</sup>. In particolare, questo corpus è stato consultato per verificare il tipo di esemplari specifici e la variabilità lessicale di ciascuna costruzione.<sup>21</sup> Si è scelto di estrarre gli esempi da tale corpus poiché, data la sua grandezza, avrebbe consentito di estrarre un numero elevato di esemplari per ciascuna costruzione. Nello specifico, i dati sono stati estratti attraverso una ricerca basata sul linguaggio CQL<sup>22</sup>; questo tipo di estrazione consente di specificare la struttura sintagmatica presa in esame, sia in termini di parte del discorso (o di elemento lessicale specifico) sia in termini di distribuzione all'interno del sintagma.

Anche per lo spagnolo e il francese sono stati utilizzati corpora interrogabili attraverso il software *SketchEngine* (il corpus *Spanish Web 2018*<sup>23</sup> e il corpus *French Web*

<sup>20</sup> Il corpus *Italian Web 2016* (*itTenTen16*) è una raccolta di dati di italiano scritto. Contiene testi estratti dal web per un totale di 4,9 miliardi di parole. Il corpus appartiene alla famiglia dei corpora *TenTen*, vaste raccolte di dati costruite secondo lo stesso metodo. I corpora *TenTen* sono consultabili attraverso il software *SketchEngine* (<https://app.sketchengine.eu/>).

<sup>21</sup> Vale la pena notare che l'intero campione di dati a cui si fa riferimento in questo contributo costituisce solo una minima parte delle possibilità combinatorie morfosintattiche e lessicali della lingua. Come è noto, i corpora sono una rappresentazione limitata della lingua, che considera esclusivamente la varietà selezionata (in questo caso, la lingua scritta del web) e un numero limitato, seppur rappresentativo, di dati.

<sup>22</sup> Il linguaggio CQL (*Corpus Query Language*) consente di estrarre pattern sintagmatici complessi, specificandone una molteplicità di caratteristiche in una singola ricerca.

<sup>23</sup> Il corpus *Spanish Web 2018* contiene 17.5 miliardi di parole estratte dal web.

2017<sup>24</sup>). Per lo spagnolo, laddove necessario, si è fatto riferimento anche al *Corpus del Español Actual*, interrogato tramite il software interno di cui è provvisto. Anche nel caso dello spagnolo e del francese, l'individuazione degli esemplari specifici di ogni costruzione è stata resa possibile attraverso la ricerca avanzata di concordanze basata sul linguaggio CQL e la relativa estrazione dei dati di frequenza.

	Italiano	Spagnolo	Francese
<b>Software usato per la raccolta delle costruzioni</b>	• EXTra	• SketchEngine • <i>Corpus del Español Actual</i>	• SketchEngine
<i>Corpus</i> usato	• <i>La Repubblica</i> • <i>Paisà</i>	• <i>Spanish Web 2018</i> (esTenTen18) • <i>Corpus del Español Actual</i>	• <i>French Web 2017</i> (frTenTen17)
<b>Software usato per la raccolta degli esempi</b>	• SketchEngine	• SketchEngine	• SketchEngine
<i>Corpus</i> usato	• <i>Italian Web 2016</i> (itTenTen16)	• <i>Spanish Web 2018</i> (esTenTen18)	• <i>French Web 2017</i> (frTenTen17)

Tabella 2. Elenco degli strumenti utilizzati per la raccolta degli esempi

In tutti i casi, per l'individuazione di sintagmi, sono anche state effettuate ricerche basate sulla funzione *n-grams* di cui è dotato il software *SketchEngine*. Tale funzione consente di identificare le sequenze contigue di unità linguistiche che occorrono con maggiore frequenza in un *corpus*; per la ricerca è necessario specificare esclusivamente il numero di posizioni (da 1 a 6) da rintracciare, senza doverne precisare il tipo morfosintattico. Tale approccio statistico permette di individuare in maniera piuttosto rapida e precisa le sequenze di parole (o, meglio, di classi di parole) che con più alta probabilità tendono a combinarsi tra loro in una certa lingua.

In totale sono stati raccolti circa 50 formati parzialmente riempiti delle diverse lingue, molti dei quali si sovrappongono, data la comune origine romanza. I formati parzialmente riempiti raccolti sono stati successivamente classificati sulla base di diversi parametri (es. proprietà sintattiche, semantiche, grado di specificità lessicale e di produttività, livello di idiomatilità), al fine di produrre una possibile rappresentazione formale che tenesse conto della comparazione interlinguistica.

#### 4. Risultati

In letteratura, i diversi tipi combinatori sono generalmente distinti sulla base delle restrizioni a cui sono soggetti (di natura morfologica, sintattica, semantica, lessicale o fonologica),<sup>25</sup> che mirano a misurarne il grado di fissità sintattica, semantica

<sup>24</sup> Il *corpus French Web 2017* contiene 5,7 miliardi di parole estratte dal web.

<sup>25</sup> Il grado di coesione e di lessicalizzazione è in genere misurato sulla base dell'applicabilità di alcuni test di tipo morfologico (per es. pluralizzazione dei costituenti), sintattico (per es. inversione dell'ordine dei costituenti, separabilità dei costituenti, ellissi di un costituente), semantico (per es. idiomatilità ed eventuale perdita del significato compositivo), lessicale (per es. variabilità lessicale o presenza di arcaismi), fonologico (per es. perdita di autonomia accentuale).

e lessicale<sup>26</sup> e il livello di lessicalizzazione. Dal punto di vista strutturale, inoltre, le diverse sequenze combinatorie sono di norma distinte in base al formato sintagmatico che le rappresenta (vale a dire la specifica sequenza di parti del discorso) (cfr. Voghera 1994, 2004). Nell'ambito della presente analisi, tuttavia, si è scelto di individuare dei criteri specifici per mezzo dei quali è possibile identificare le CPR e distinguerle dagli altri fenomeni combinatori. In particolare, la classificazione riportata nell'ambito del presente contributo è essenzialmente basata su i) criteri sintattici, ii) criteri semantici e iii) il livello di produttività delle diverse configurazioni<sup>27</sup>. Vale la pena notare che i dati analizzati in questo contributo includono perlopiù strutture che trovano piena corrispondenza in tutte le lingue considerate: alcune configurazioni possono infatti ricorrere nella medesima funzione e struttura in più lingue della stessa famiglia (cfr. la famiglia romanza negli esempi considerati).

#### 4.1. Criteri sintattici

Dal punto di vista sintattico, possono essere presi in considerazione più criteri per la classificazione delle CPR. Da un lato, può essere utile differenziare la tipologia costruzionale sulla base delle unità sintattiche attraverso le quali tali costruzioni si manifestano (es. sintagmi o strutture frasali). Dall'altro lato, è importante tenere in considerazione le proprietà strutturali interne al sintagma o alla frase, ovvero il numero di elementi di cui si compongono e, in particolare, il numero e il tipo di posizioni soggette a variazione.

Per quanto riguarda il primo punto, le strutture parzialmente riempite possono associarsi a diversi tipi di costruzioni, che possono operare sia a livello frasale sia a livello sintagmatico. Appartiene alle strutture frasali, ad esempio, la costruzione *let alone*, analizzata in Fillmore *et al.* (1988):

- (2) a. His parents let him drink<sub>F</sub> whiskey<sub>A</sub>, let alone coke<sub>B</sub>  
b. He doesn't get up for<sub>F</sub> lunch<sub>A</sub>, let alone breakfast<sub>B</sub>

In questo caso la costruzione presenta un'unica posizione fissa e lessicalmente specificata (quella ricoperta da *let alone*) e da alcune posizioni vuote (i 'frames sintattici' F, A, B), all'interno delle quali possono essere introdotti diversi elementi frasali o lessicali. L'intera struttura ha una specifica semantica e una peculiare funzione pragmatica che non possono essere ricondotte alla somma dei costituenti.

<sup>26</sup> Anche denominato *figement* (Gross 1996, Meiri 1997, Lamiroy 2003) o *forza coesiva* (Simone 2007), tale criterio coincide con la cristallizzazione dell'unità combinatoria. In letteratura, il *figement* è tradizionalmente distinto in i) sintattico (se concerne i diversi costituenti del sintagma), e ii) semantico (se legato all'opacità semantica e idiomacità della sequenza combinatoria). In genere, le costruzioni mostrano un *figement* di tipo sia sintattico sia semantico.

<sup>27</sup> Come è noto, la produttività delle configurazioni può variare con il tempo. Se, da un lato, la schematicità di un formato multiparola può aumentare e rimanere viva nelle fasi più recenti di lingua, dall'altro lato alcune strutture possono essere molto produttive nelle fasi di lingua più antica, e presentare una bassa (o inesistente) produttività nelle fasi successive. Tale tematica non sarà tuttavia trattata nell'ambito del presente lavoro. Si rimanda a Piunno - Ganfi (in stampa) per approfondimenti.

Le costruzioni prese in considerazione per il presente studio sono di natura sintagmatica, vale a dire non superano di norma l'estensione di un sintagma. Dal punto di vista della categoria sintagmatica, si può notare che le CPR possono ricoprire la stessa funzione delle parole contenute. Tali sequenze possono pertanto veicolare dei significati pieni, come nel caso del sintagma nominale (3) o del sintagma verbale (4)). Gli esempi che seguono riportano anche i più frequenti *fillers* (gli elementi lessicali a riempimento delle posizioni vuote) registrati nel *corpus* (cfr. tabelle 3-4-5), la loro frequenza di occorrenza ("fq") e la loro frequenza normalizzata sulla base della grandezza del *corpus*:

(3) Italiano<sup>28</sup>

- FORMATO: *colpo di* X<sub>NOME <strumento></sub><sup>29</sup>
- SIGNIFICATO: i) 'movimento o azione rapida svolta attraverso X<sub>NOME <strumento></sub>'  
ii) 'esplosione provocata dalla messa in funzione di X<sub>NOME <arma da fuoco></sub>'
- ESEMPLI: a. «Dall'auto vennero esplosi colpi di pistola per intimidire...»  
b. «Con tre colpi di bacchetta, egli ritornò quello stesso che era stato prima»

Semantica i = 'Esplosione provocata dalla messa in funzione di X <sub>NOME &lt;arma da fuoco&gt;</sub> '				Semantica i = 'Movimento o azione rapida svolta attraverso X <sub>NOME &lt;strumento&gt;</sub> '			
	Sequenza (forme)	Fq	Frequenza Normalizzata		Sequenza (forme)	Fq	Frequenza Normalizzata
1.	<i>colpo di pistola</i>	10010	170.688	1.	<i>colpo di bastone</i>	598	0.10197
2.	<i>colpo di fucile</i>	3588	0.61182	2.	<i>colpo di machete</i>	508	0.08662
3.	<i>colpo di cannone</i>	1913	0.32620	3.	<i>colpo di bacchetta</i>	502	0.08560
4.	<i>colpo di rivoltella</i>	543	0.09259	4.	<i>colpo di piccone</i>	474	0.08083
5.	<i>colpo di mitra</i>	369	0.06292	5.	<i>colpo di mazza</i>	409	0.06974

Tabella 3. Primi 10 esempi nel corpus itTenTen16

Di seguito un esempio di costruzione verbale altamente produttiva (vale a dire, con un numero elevato di esemplari registrati nel *corpus*):

## (4) Italiano

- FORMATO: *avere il* X<sub>NOME <concreto/evento></sub> *facile*
- SIGNIFICATO: 'essere incline a fare qualcosa connesso con X<sub>NOME</sub>'
- ESEMPLI: a. «Lei che ha la penna facile, mi faccia un favore, mi scriva due righe per la rivista»  
b. «Mi ritengo più un giornalista che un poeta anche se ho la rima facile»

<sup>28</sup> Quello mostrato è un pattern abbastanza produttivo, presente già nelle fasi più antiche delle tre lingue romanze considerate. Per approfondimenti sul tema dei sintagmi preposizionali con funzione aggettivale o avverbale dell'italiano, si rimanda a Piunno (2018); uno studio diacronico è presente in Piunno - Ganfi (in stampa).

<sup>29</sup> Nelle rappresentazioni formali fornite nel presente studio, le parti invariabili sono segnalate in corsivo, mentre quelle variabili sono in tondo.

	Sequenza (lemma)	Fq	Frequenza Normalizzata		Sequenza (lemma)	Fq	Frequenza Normalizzata
1.	<i>avere la lacrima facile</i>	98	0.01671	21.	<i>avere la manetta facile</i>	2	0.00034
2.	<i>avere il grilletto facile</i>	34	0.00580	22.	<i>avere la domanda facile</i>	2	0.00034
3.	<i>avere la querela facile</i>	18	0.00307	23.	<i>avere la chiacchiera facile</i>	2	0.00034
4.	<i>avere la penna facile</i>	16	0.00273	24.	<i>avere il whisky facile</i>	2	0.00034
5.	<i>avere la denuncia facile</i>	12	0.00205	25.	<i>avere il vomito facile</i>	2	0.00034
6.	<i>avere il pianto facile</i>	10	0.00171	26.	<i>avere il tradimento facile</i>	2	0.00034
7.	<i>avere la lacrimuccia facile</i>	9	0.00153	27.	<i>avere il rosso facile</i>	2	0.00034
8.	<i>avere il cartellino facile</i>	8	0.00136	28.	<i>avere il riso facile</i>	2	0.00034
9.	<i>avere la risata facile</i>	7	0.00119	29.	<i>avere il perdono facile</i>	2	0.00034
10.	<i>avere il sonno facile</i>	7	0.00119	30.	<i>avere il manganello facile</i>	2	0.00034
11.	<i>avere la battuta facile</i>	6	0.00102	31.	<i>avere il ditino facile</i>	2	0.00034
12.	<i>avere la pistola facile</i>	5	0.00085	32.	<i>avere il coltello facile</i>	2	0.00034
13.	<i>avere la commozione facile</i>	4	0.00068	33.	<i>avere la vendetta facile</i>	1	0.00017
14.	<i>avere il gol facile</i>	4	0.00068	34.	<i>avere la tintarella facile</i>	1	0.00017
15.	<i>avere la uccisione facile</i>	3	0.00051	35.	<i>avere la stecca facile</i>	1	0.00017
16.	<i>avere la risposta facile</i>	3	0.00051	36.	<i>avere la spada facile</i>	1	0.00017
17.	<i>avere la menzogna facile</i>	3	0.00051	37.	<i>avere la sforbiciata facile</i>	1	0.00017
18.	<i>avere il sorriso facile</i>	3	0.00051	38.	<i>avere la sbronza facile</i>	1	0.00017
19.	<i>avere la rima facile</i>	2	0.00034	39.	<i>avere la rissa facile</i>	1	0.00017
20.	<i>avere la mano facile</i>	2	0.00034	40.	<i>avere la provocazione facile</i>	1	0.00017

Tabella 4. Primi 40 esempi nel corpus itTenTen16

Alcune sequenze possono avere la forma di un sintagma preposizionale, che, in funzione aggettivale, può legarsi ad un nome (cfr. Piunno 2018):

(5) Italiano<sup>30</sup>

- FORMATO:  $X_{\text{NOME} \langle \text{concreto} \rangle} \textit{alla} \text{AGG}_{\langle \text{femminile} \rangle}$
- SIGNIFICATO:  $X_{\text{NOME} \langle \text{concreto} \rangle} \textit{fatto alla maniera tipica di} Y_{\text{AGG}}$
- ESEMPIO: «Se siete appassionati del *poker alla texana* e volete diventare dei veri professionisti...»

	Sequenza (forme)	Fq	Frequenza Normalizzata		Sequenza (forme)	Fq	Frequenza Normalizzata
1.	pesto <i>alla</i> genovese	924	0.15756	6.	pantaloni <i>alla</i> zuava	198	0.03376
2.	risotto <i>alla</i> milanese	790	0.13471	7.	poker <i>alla</i> texana	165	0.02814
3.	spaghetti <i>alla</i> carbonara	553	0.09430	8.	pesto <i>alla</i> trapanese	157	0.02677
4.	baccalà <i>alla</i> vicentina	368	0.06275	9.	ragù <i>alla</i> bolognese	114	0.01944
5.	riso <i>alla</i> cantonese	219	0.03734	10.	aragosta <i>alla</i> catalana	78	0.01330

Tabella 5. Primi 10 esempi nel corpus itTenTen16

Vale la pena notare che gli aggettivi sostantivati impiegati all'interno di tali sequenze rientrano all'interno di una scelta paradigmatica potenzialmente infinita, tra cui è possibile identificare in particolare gli aggettivi etnici. Dal punto di vista morfosintattico, la costruzione prevede che l'aggettivo sia sempre flesso al femminile (cfr. Piunno 2018).

Le CPR possono anche dare luogo a parole funzione e svolgere, pertanto, una funzione grammaticale, come nel caso dei sintagmi preposizionali complessi dell'italiano (6) (cfr. Ganfi – Piunno 2017, Piunno – Ganfi 2019).

## (6) Italiano

- FORMATO:  $su X_{\text{NOME} \langle \text{evento} \rangle} di^{31}$
- SIGNIFICATO: 'effettuato a seguito di un/a  $X_{\text{NOME} \langle \text{evento} \rangle}$  da parte di'
- ESEMPI:
  - a. «Nel 1923, *su invito di* un sacerdote, scrisse e interpretò una commedia»
  - b. «Rachel riprogettò *su incarico di* Jacqueline Kennedy i giardini della Casa Bianca»

<sup>30</sup> Quello mostrato è un pattern abbastanza produttivo, presente già nelle fasi più antiche delle tre lingue romanze considerate. Per approfondimenti sul tema dei sintagmi preposizionali con funzione aggettivale o avverbale dell'italiano, si rimanda a Piunno (2018); uno studio diacronico è presente in Piunno – Ganfi (in stampa).

<sup>31</sup> Come notato in Piunno – Ganfi (2019), il nome eventivo incluso nella costruzione proviene spesso da una nominalizzazione transitiva.

	Sequenza (lemma)	Fq	Frequenza Normalizzata		Sequenza (lemma)	Fq	Frequenza Normalizzata
1.	<i>su richiesta di</i>	7617	129.883	11.	<i>su sollecitazione di</i>	805	0.13727
2.	<i>su iniziativa di</i>	4928	0.84031	12.	<i>su presentazione di</i>	720	0.12277
3.	<i>su proposta di</i>	2389	0.40737	13.	<i>su mandato di</i>	533	0.09089
4.	<i>su consiglio di</i>	2265	0.38622	14.	<i>su intervento di</i>	408	0.06957
5.	<i>su invito di</i>	2073	0.35348	15.	<i>su pressione di</i>	380	0.06480
6.	<i>su suggerimento di</i>	1695	0.28903	16.	<i>su domanda di</i>	312	0.05320
7.	<i>su segnalazione di</i>	1543	0.26311	17.	<i>su disposizione di</i>	180	0.03069
8.	<i>su commissione di</i>	1314	0.22406	18.	<i>su denuncia di</i>	179	0.03052
9.	<i>su incarico di</i>	1281	0.21843	19.	<i>su delega di</i>	173	0.02950
10.	<i>su indicazione di</i>	1199	0.20445	20.	<i>su valutazione di</i>	160	0.02728

Tabella 6. Primi 20 esempi nel corpus itTenTen16

Infine, vale la pena notare che, a differenza di quelle frasali, le CPR di natura sintagmatica non hanno, di norma, una funzione pragmatica. Possono, tuttavia, non essere sempre prive di tratti di natura connotativa (come nell'es. (4), in cui la costruzione richiama una valutazione – prevalentemente – negativa del parlante legata all'eccessivo uso che si fa di un certo nome concreto X).

Dal punto di vista della struttura interna del sintagma, le CPR possono essere caratterizzate dalla presenza di una o più posizioni soggette a variazione. La struttura più frequente è quella in cui la costruzione possiede un solo *slot* lessicalmente variabile, come nell'esempio che segue, tratto dallo spagnolo:

## (7) Spagnolo

·FORMATO: *dar por* X<sub>AGG/PART.PRES./PART.PASS.</sub>·SIGNIFICATO: 'considerare come X<sub>AGG/PART.PRES./PART.PASS.</sub>'·ESEMPLI: a. «Sólo Filósofos muy alucinados *darán por buena* tal ilación»b. «Tampoco *deis por verdadera* cualquier cosa que encontréis por internet»

Nel caso esemplificato in (7) si ha una singola posizione sintattica aperta, che mostra una variabilità lessicale piuttosto ampia. Può essere infatti saturata da una tipologia molto differenziata di aggettivi o participi (presenti o passati), non sottoposti a una specifica restrizione semantica, come mostrato nella tabella 7.

La costruzione dello spagnolo trova piena corrispondenza in francese e in italiano.

(8) Italiano a. *Ma ho dato per buono* che ne sapessero più di me(9) Francese b. Il y a quelques semaines, la nomination était *donnée pour imminente* et décidée

Si osservi, a tal proposito, la tabella 8.

	Sequenza (lemma)	Fq	Frequenza Normalizzata		Sequenza (lemma)	Fq	Frequenza Normalizzata
1.	<i>dar por</i> bueno 'dare per buono'	15620	0.76921	11.	<i>dar por</i> existente 'considerare esistente'	197	0.00970
2.	<i>dar por</i> válido 'dare per valido'	4515	0.22234	12.	<i>dar por</i> extinto 'considerare estinto'	169	0.00832
3.	<i>dar por</i> seguro 'considerare sicuro'	2780	0.13690	13.	<i>dar por</i> falso 'considerare falso'	157	0.00773
4.	<i>dar por</i> cierto 'dare per certo'	1973	0.09716	14.	<i>dar por</i> evidente 'considerare evidente'	153	0.00753
5.	<i>dar por</i> imposible 'dare per impossibile'	935	0.04604	15.	<i>dar por</i> concluso 'considerare concluso'	145	0.00714
6.	<i>dar por</i> completo 'considerare completo'	606	0.02984	16.	<i>dar por</i> oficial 'considerare ufficiale'	119	0.00586
7.	<i>dar por</i> verdadero 'considerare vero'	401	0.01975	17.	<i>dar por</i> inminente 'dare per imminente'	42	0.00207
8.	<i>dar por</i> definitivo 'considerare definitivo'	382	0.01881	18.	<i>dar por</i> probable 'considerare probabile'	33	0.00163
9.	<i>dar por</i> obvio 'considerare ovvio'	356	0.01753	19.	<i>dar por</i> irrecuperable 'considerare irrecuperabile'	29	0.00143
10.	<i>dar por</i> correcto 'considerare corretto'	305	0.01502	20.	<i>dar por</i> efectivo 'considerare effettivo'	27	0.00133

Tabella 7. Primi 20 esempi nel corpus esTenTen18

È interessante notare, tuttavia, alcune differenze tra le lingue prese in considerazione. In primo luogo, in spagnolo e in italiano la costruzione mostra una buona produttività in termini di numero di forme<sup>32</sup> e un'alta frequenza di occorrenza. Più è alto il numero di forme e la frequenza d'uso di ciascuna forma, più è probabile che la costruzione a cui si riconduce tale forma abbia acquisito dei tratti di schematicità; ad un alto grado di schematicità si associa di norma un alto livello di produttività (Barðdal 2008: 45). I dati del francese si mostrano notevolmente diversi. In francese, infatti, sia il numero di strutture identificate<sup>33</sup> sia la relativa frequenza d'uso

<sup>32</sup> Le strutture estratte per lo spagnolo sono 1001, mentre per l'italiano sono 649. Nonostante i dati numerici assoluti non siano comparabili tra loro, le frequenze normalizzate ci dicono la loro importanza quantitativa in relazione ai rispettivi corpora.

<sup>33</sup> Le strutture contenenti la sequenza sintattica in oggetto estratte per il francese sono 232. Tra queste, tuttavia, si registrano molti esempi la cui semantica non corrisponde a quella evidenziata in (7) per lo spagnolo.



(si noti in particolare la frequenza normalizzata) sono estremamente ridotte. In secondo luogo, può accadere che, pur essendo aperte allo stesso tipo di variazione paradigmatica e pur veicolando il medesimo significato, alcune costruzioni siano sottoposte a restrizioni lessicali specifiche delle diverse lingue. Nel caso specifico, ad esempio, la costruzione italiana include frequentemente il lessema *favorito* (nel corpus ITTenTen16, l'aggettivo/participio passato *favorito* è tra i primi 20 più frequentemente associati a questa struttura, su un totale di circa 900 lemmi), mentre in spagnolo l'uso di questo lessema ha una frequenza estremamente bassa. Inoltre, vale la pena notare che gli equivalenti di *dar por* in italiano possono essere espressi attraverso due strutture diverse (*dare per* + agg./part. o *considerare* + agg./part.), non sempre intercambiabili tra loro<sup>34</sup>.

ITALIANO				FRANCESE			
	Sequenza (lemma)	Fq	Frequenza Normalizzata		Sequenza (lemma)	Fq	Frequenza Normalizzata
1.	<i>dare per scontato</i>	22524	384.074	1.	<i>donner pour mort</i> 'dare per morto'	302	0.04412
2.	<i>dare per certo</i>	1438	0.24520	2.	<i>donner pour vrai</i> 'dare per vero	57	0.00833
3.	<i>dare per vinto</i>	1284	0.21894	3.	<i>donner pour certain</i> 'dare per certo'	27	0.00394
4.	<i>dare per acquisito</i>	524	0.08935	4.	<i>donner pour sûr</i> 'dare per sicuro'	11	0.00161
5.	<i>dare per buono</i>	409	0.06974	5.	<i>donner pour équivalent</i> 'dare per equivalente'	22	0.00321
6.	<i>dare per primo</i>	332	0.05661	6.	<i>donner pour évident</i> 'dare per evidente'	8	0.00117
7.	<i>dare per imminente</i>	277	0.04723	7.	<i>donner pour imminent</i> 'dare per imminente'	8	0.00117
8.	<i>dare per inteso</i>	250	0.04263	8.	<i>donner pour bon</i> 'dare per buono'	7	0.00102
9.	<i>dare per disperso</i>	243	0.04144	9.	<i>donner pour probable</i> 'dare per probabile'	4	0.00058
10.	<i>dare per perso</i>	192	0.03274	10.	<i>donner pour impossible</i> 'dare per impossibile'	3	0.00044

Tabella 8. Primi 10 esempi nel corpus itTenTen16 (italiano) e nel corpus frTenTen17 (francese)

Alcune strutture si presentano come costruzioni a più posizioni vuote. In questi casi, le porzioni di sintagma sottoposte a variazione possono essere completate attraverso una gamma più o meno ampia di elementi lessicali. Si osservino a tal proposito gli esempi che seguono, dal francese:

(10) Francese

·FORMATO: [[*de* + X<sub>N1<tempo></sub>] [*en* + X<sub>N1<tempo></sub>]]

·SIGNIFICATO: 'progressione temporale'

·ESEMPIO: «Ma vue s'améliore *de jour en jour*»

<sup>34</sup> Per stabilire con precisione le proprietà lessico-semantiche di queste costruzioni sarebbe necessario prendere in considerazione un più ampio numero di dati e di configurazioni sintagmatiche. Tale obiettivo va, tuttavia, oltre gli scopi del presente lavoro.

	Sequenza (lemma)	Fq	Frequenza Normalizzata
1.	<i>de temps en temps</i> 'di tanto in tanto'	74085	1.082.223
2.	<i>de jour en jour</i> 'di giorno in giorno'	11299	165.054
3.	<i>de année en année</i> 'di anno in anno'	11239	164.178
4.	<i>de heure en heure</i> 'di ora in ora'	1048	0.15309
5.	<i>de moi en moi</i> 'di mese in mese'	1026	0.14988
6.	<i>de siècle en siècle</i> 'di secolo in secolo'	662	0.09670
7.	<i>de âge en âge</i> 'di epoca in epoca'	624	0.09115
8.	<i>de minute en minute</i> 'di minuto in minuto'	483	0.07056
9.	<i>de instant en instant</i> 'di attimo in attimo'	348	0.05084
10.	<i>de second en second</i> 'di secondo in secondo'	155	0.02000

Tabella 9. Primi 10 esempi nel corpus frTenTen17

In questo caso la porzione variabile può essere saturata da i) un unico elemento lessicale che si ripete nelle due diverse posizioni, che deve necessariamente ii) appartenere alla classe nominale, e iii) indicare un'unità temporale. Considerate le diverse restrizioni sintattico-lessico-semantiche, la costruzione acquisisce un valore di 'progressione temporale'. Inoltre, vale la pena notare che la struttura presenta un numero estremamente limitato di *fillers* e di esemplari, e quindi una bassa produttività. Alcuni esemplari registrano una frequenza notevolmente alta rispetto ad altri (si osservino a tal proposito i primi tre esempi): si tratta di sequenze che hanno raggiunto, col tempo, alti gradi di coesione e lessicalizzazione.<sup>35</sup>

Come evidenziato in Piunno (2018), alcune costruzioni possiedono una struttura ancor meno specificata a livello lessicale. Si tratta di sequenze dalla natura più astratta o schematica, che prevedono almeno due posizioni vuote, che sono sottoposte a restrizioni semantico/lessicali di natura diversa, e un'unica posizione riempita. Si osservino i seguenti esempi, tratti dal francese:

(11) Francese

- FORMATO: [N<sub>1</sub><concreto> *à/de* N<sub>2</sub><forza naturale/materiale/strumento/parte del corpo/evento>]
- SIGNIFICATO: 'X<sub>N1</sub><strumento> che funziona mediante Y<sub>N2</sub><forza naturale/materiale/strumento/parte del corpo/evento>']
- ESEMPIO: «Le château est construit à partir d'un ancien *moulin à eau*»

<sup>35</sup> A tal proposito, vale la pena notare che, sul piano diacronico, le CPR possono variare sotto diversi aspetti, come ad esempio i) il livello di schematicità, ii) la produttività e iii) l'idiomaticità. Un'analisi delle CPR che tenga in considerazione la loro evoluzione nel tempo consente di capire come alcuni schemi, attraverso la frequenza d'uso, abbiano aumentato o diminuito il grado di produttività. Inoltre, lo studio delle diverse fasi di lingua può essere utile a rintracciare i possibili percorsi attraverso i quali sono stati generate alcune sequenze sintagmatiche più recenti, fortemente idiosincratice e dal significato altamente idiomatico (cfr. Piunno - Ganfi, in stampa), che appaiono altamente lessicalizzate negli stadi di lingua più recenti.

	Sequenza (lemma)	Fq	Frequenza Normalizzata
1.	arme à feu 'arma da fuoco'	7831	114.394
2.	pompe à chaleur 'pompa a calore'	5815	0.84945
3.	moulin à vent 'mulino a vento'	3381	0.49389
4.	machine à vapeur 'macchina a vapore'	3321	0.48513
5.	moulin à eau 'mulino ad acqua'	2389	0.34898
6.	instrument à vent 'strumento a vento'	1857	0.27127
7.	véhicule à moteur 'veicolo a motore'	1856	0.27112
8.	planche à voile 'tavola da surf' (lett. 'a vela')	1812	0.26469
9.	moteur à explosion 'motore a scoppio'	705	0.10299
10.	frein à main 'freno a mano'	621	0.09071

Tabella 10. Primi 10 esempi nel corpus frTenTen17

Nell'esempio del francese (che trova piena corrispondenza anche nelle altre lingue romanze, cfr. Piunno 2018), si può notare che, nonostante entrambe le posizioni vuote richiedano degli elementi nominali ( $X_{N1}$  e  $Y_{N2}$ ), i due slot sono tuttavia caratterizzati da proprietà semantiche differenti (la posizione  $X_{N1}$  descrive uno strumento, rappresentato da un nome di oggetto concreto, mentre  $Y_{N2}$  descrive il modo in cui  $X_{N1}$  acquisisce la possibilità di funzionare). In questi casi, pur avendo delle diverse possibilità combinatorie e pur mostrando un'ampia apertura a range paradigmatici di diversa natura (in particolare per  $Y_{N2}$ ), la costruzione mantiene la semantica univoca di 'che funziona per mezzo di'.

#### 4.2. Criteri semantici

Dal punto di vista semantico, le CPR possono essere analizzate secondo diversi parametri, tra cui ad esempio:

- i. presenza di restrizioni semantiche specifiche;
- ii. presenza di una semantica *costruzionale*;
- iii. interazione tra restrizioni semantiche e significato costruzionale.

Il primo criterio concerne le restrizioni semantiche a cui alcune costruzioni, e in particolare alcune porzioni di esse, possono essere sottoposte. In molti casi, infatti, la selezione degli elementi lessicali da usare per la saturazione delle posizioni vuote è condizionata da specifiche restrizioni, legate alla natura semantica del lessema (o del gruppo di lessemi) che può ricoprire tale posizione. Si osservi, a tal proposito, la seguente CPR:

(12) Italiano

- FORMATO:  $X_{NOME <concreto>}$  *alla mano*
- SIGNIFICATO:  $X_{NOME <concreto>}$  'pronto da usare'
- ESEMPI:
  - a. «Questo libro dimostra, *documenti alla mano*, che proprio dove i controlli dovrebbero essere garantiti, c'è il massimo dell'opacità»
  - b. «*Conti alla mano*, sono convinto si possa già fare tutto oggi senza aspettare il 2018»

	Sequenza (forme)	Fq	Frequenza Normalizzata		Sequenza (forme)	Fq	Frequenza Normalizzata
1.	dati <i>alla mano</i>	4280	0.72982	21.	libro <i>alla mano</i>	99	0.01688
2.	numeri <i>alla mano</i>	1423	0.24265	22.	mappa <i>alla mano</i>	95	0.01620
3.	conti <i>alla mano</i>	950	0.16199	23.	guida <i>alla mano</i>	87	0.01484
4.	documenti <i>alla mano</i>	685	0.11680	24.	coltello <i>alla mano</i>	83	0.01415
5.	carte <i>alla mano</i>	672	0.11459	25.	bici <i>alla mano</i>	83	0.01415
6.	statistiche <i>alla mano</i>	492	0.08389	26.	i-pad <i>alla mano</i>	81	0.01381
7.	cifre <i>alla mano</i>	447	0.07622	27.	metro <i>alla mano</i>	77	0.01313
8.	armi <i>alla mano</i>	446	0.07605	28.	codice <i>alla mano</i>	77	0.01313
9.	calcolatrice <i>alla mano</i>	270	0.04604	29.	microfono <i>alla mano</i>	76	0.01296
10.	prove <i>alla mano</i>	269	0.04587	30.	spada <i>alla mano</i>	71	0.01211
11.	calendario <i>alla mano</i>	257	0.04382	31.	costituzione <i>alla mano</i>	64	0.01091
12.	cartina <i>alla mano</i>	250	0.04263	32.	soldi <i>alla mano</i>	62	0.01057
13.	cronometro <i>alla mano</i>	200	0.03410	33.	fucile <i>alla mano</i>	62	0.01057
14.	risultati <i>alla mano</i>	196	0.03342	34.	taccuino <i>alla mano</i>	59	0.01006
15.	classifica <i>alla mano</i>	186	0.03172	35.	legge <i>alla mano</i>	59	0.01006
16.	pistola <i>alla mano</i>	158	0.02694	36.	bilancio <i>alla mano</i>	57	0.00972
17.	regolamento <i>alla mano</i>	134	0.02285	37.	atto <i>alla mano</i>	57	0.00972
18.	penna <i>alla mano</i>	129	0.02200	38.	chitarra <i>alla mano</i>	54	0.00921
19.	calcoli <i>alla mano</i>	102	0.01739	39.	telecamera <i>alla mano</i>	50	0.00853
20.	sondaggi <i>alla mano</i>	100	0.01705	40.	progetto <i>alla mano</i>	49	0.00836

Tabella 11. Primi 40 esempi nel corpus itTenTen16

La costruzione in oggetto registra elevati livelli di produttività e schematicità, determinati i) dall'elevato numero di esemplari rintracciabili nel corpus e ii) dalla loro bassa frequenza di occorrenza: ad un alto numero di *fillers* con frequenza bassa può essere infatti associato un alto grado di produttività e astrattezza della costruzione (cfr. Bybee 2006, 2010; Barðdal 2008; Ziem 2018). Nel caso citato, specifico dell'italiano, il sintagma preposizionale *alla mano* – un tipico esempio delle strutture che in Piunno (2018) vengono definite *aggettivali sintagmatici* – acquisisce il significato peculiare di 'pronto da usare' solo se preceduto da nomi che denotano oggetti concreti, strumenti; in questo caso l'aggettivale, unica porzione fissa del sintagma, identifica una sorta di relazione spaziale tra lo strumento denotato da  $X_{N1}$  e il partecipante che,

attraverso la *mano*, lo impugna. La semantica dell'aggettivale *alla mano* è dunque strettamente condizionata dalla restrizione semantica su  $N_1$ , legata alla concretezza. Ne è una prova il fatto che se si sostituisce il nome concreto con un nome di individuo (per es. *persona alla mano*), l'aggettivale cambia drasticamente significato in 'affabile, cordiale'.<sup>36</sup> Le strutture sintatticamente complesse possono presentare un grado di composizionalità semantica variabile, e talvolta una semantica costruzionale specifica.<sup>37</sup> Questo significa che gli elementi che compongono la sequenza possono perdere una parte del loro significato e acquisire un valore specifico nel contesto costruzionale, come accade negli esempi precedenti. In particolare, in taluni casi, a variare è la sola semantica di alcune parti del sintagma complesso (cfr. *alla mano*, nei due valori individuati), sulla base della presenza di specifici elementi lessicali. In altri casi, la semantica dell'intera sequenza costruzionale può subire una lieve variazione a seconda degli elementi lessicali selezionati, come negli esempi che seguono:

(13) Italiano

FORMATO: [darsi a + Det +  $X_{N1}$ <attività>]

SIGNIFICATO: 'dedicarsi a  $X_{N1}$ '

ESEMPIO: «Appena abbiamo visto i prezzi americani *ci siamo dati allo shopping*»

	Sequenza (lemma)	Fq	Frequenza Normalizzata		Sequenza (lemma)	Fq	Frequenza Normalizzata
1.	<i>darsi allo studio</i>	107	0.01825	11.	<i>darsi allo shopping</i>	28	0.00477
2.	<i>darsi alla politica</i>	93	0.01586	12.	<i>darsi al cinema</i>	26	0.00443
3.	<i>darsi al saccheggio</i>	86	0.01466	13.	<i>darsi alla pirateria</i>	20	0.00341
4.	<i>darsi al commercio</i>	47	0.00801	14.	<i>darsi alla letteratura</i>	20	0.00341
5.	<i>darsi alla ricerca</i>	45	0.00767	15.	<i>darsi alla lettura</i>	19	0.00324
6.	<i>darsi alla musica</i>	36	0.00614	16.	<i>darsi al lavoro</i>	19	0.00324
7.	<i>darsi alla carriera</i>	35	0.00597	17.	<i>darsi al gioco</i>	19	0.00324
8.	<i>darsi alla caccia</i>	35	0.00597	18.	<i>darsi alla pittura</i>	18	0.00307
9.	<i>darsi alla prostituzione</i>	30	0.00512	19.	<i>darsi al sesso</i>	17	0.00290
10.	<i>darsi al brigantaggio</i>	29	0.00495	20.	<i>darsi alla preghiera</i>	15	0.00256

Tabella 12. Primi 20 esempi nel corpus itTenTen16<sup>38</sup>

<sup>36</sup> Si potrebbe tuttavia notare che i due valori attribuibili al sintagma *alla mano* (cfr. 'pronto da usare' e 'cordiale, affabile') condividono il medesimo significato sottostante di 'facile a essere avvicinato'. I due significati, pur avendo un sostrato semantico comune, si sono differenziati (o, meglio, specializzati) diacronicamente in maniera sostanziale sia in termini di specificità semantica sia dal punto di vista delle proprietà combinatorie. Questo va a confermare il principio di non sinonimia presentato in Goldberg (1995).

<sup>37</sup> Cfr. Piunno (2015, 2018).

<sup>38</sup> Si può notare che in alcuni casi, i diversi *fillers* della struttura possono attribuire una connotazione negativa all'intera costruzione, come nel caso di *gioco*, rispetto agli altri esempi presentati.

## (14) Italiano

- FORMATO:  $[darsi\ a + Det + X_{N1<concreto>}]$   
 ·SIGNIFICATO: 'dedicarsi a/abusare di  $X_{N1}$ ' (spesso con connotazione negativa)  
 ·ESEMPIO: «Il 42% dei ragazzi e il 21% delle ragazze minorenni *si dà alla bottiglia*»

	Sequenza (lemma)	Fq	Frequenza Normalizzata
1.	<i>darsi all' alcol</i>	29	0.00495
2.	<i>darsi al vino</i>	5	0.00085
3.	<i>darsi al cibo</i>	4	0.00068
4.	<i>darsi alla bottiglia</i>	3	0.00051
5.	<i>darsi al fumo</i>	2	0.00034

Tabella 13. Primi 5 esempi nel corpus itTenTen17

Entrambi gli esempi si associano alla costruzione  $[darsi\ a + X_{N1}]$ , in cui il verbo riflessivo *darsi* è seguito dalla preposizione articolata e da un nome. Come riportato nella descrizione del formato di parola, le due costruzioni possono variare sensibilmente il loro valore semantico in base al tipo di nome selezionato. In genere la costruzione include un nome eventivo, e la sua semantica corrisponde solitamente a 'dedicarsi a una qualche attività connessa con  $X_{N1}$ ' (13). La costruzione può assumere la semantica peculiare di 'lasciarsi prendere la mano da, abusare di' – e una connotazione tipicamente negativa – quando incorpora un nome concreto: in questo caso l'oggetto indica, per mezzo di uno slittamento metonimico, la tipica attività connessa con esso (14).

Questo significa che le restrizioni semantiche applicate ai singoli elementi della costruzione possono riflettersi sul piano della semantica costruzionale (cfr. criterio iii). Di conseguenza, ad una variazione del range paradigmatico può spesso corrispondere una variazione nella semantica complessiva della costruzione. In questi casi, si ha perlopiù a che fare con sequenze costruzionali dall'alto grado di idiomaticità. Si osservino gli esempi seguenti:

(15) Spagnolo<sup>39</sup>

- FORMATO:  $[[de + X_{N1<tempo>}] [en + X_{N1<tempo>}]]$   
 ·SIGNIFICATO: 'progressione temporale'  
 ·ESEMPIO: «La marcha del burro se había hecho *de minuto en minuto*, más difícil»

	Sequenza (lemma)	Fq	Frequenza Normalizzata
1.	<i>de año en año</i> 'di anno in anno'	3731	0.18373
2.	<i>de tiempo en tiempo</i> 'di tanto in tanto'	3466	0.17068
3.	<i>de vez en vez</i> 'di tanto in tanto'	3383	0.16660
4.	<i>de día en día</i> 'di giorno in giorno'	2081	0.10248
5.	<i>de minuto en minuto</i> 'di minuto in minuto'	142	0.01

Tabella 14. Primi 5 esempi nel corpus esTenTen17

<sup>39</sup> Cfr. esempio (7) del francese.

(16) Spagnolo

- FORMATO:  $[[de + X_{N1<luogo>}] [en + X_{N1<luogo>}]]$
- SIGNIFICATO: 'progressione spaziale'
- ESEMPIO: «Iba de casa en casa y nadie le negaba nada»

	Sequenza (lemma)	Fq	Frequenza Normalizzata
1.	<i>de casa en casa</i> 'di casa in casa'	4389	0.21614
2.	<i>de pueblo en pueblo</i> 'di paese in paese'	2885	0.14207
3.	<i>de ciudad en ciudad</i> 'di città in città'	2166	0.10666
4.	<i>de país en país</i> 'di paese in paese'	1568	0.07722
5.	<i>de isla en isla</i> 'di isola in isola'	861	0.04240

Tabella 15. Primi 5 esempi nel corpus esTenTen17

Le strutture in (15)-(16) sono identiche da un punto di vista sintagmatico. Sul piano semantico, entrambe le costruzioni sono utilizzate per designare 'una progressione, un avanzamento', significato generico veicolato dalla struttura più astratta  $[[de + X_{N1}] [en + X_{N1}]]$ . La struttura più astratta può essere, pertanto, usata per designare una progressione in termini spaziali o temporali. I due diversi valori sono attribuiti sulla base della selezione lessicale operata: i) quando ingloba elementi lessicali che designano luoghi, la struttura intera acquisisce un valore spaziale, mentre ii) assume un valore temporale quando include nomi di unità di tempo.

Gli esempi trovano piena corrispondenza in termini di proprietà semantiche e sintattiche in tutte le lingue considerate.

(17) Semantica i = 'progressione temporale' Semantica ii = 'progressione spaziale'

- Francese:  $[[de + X_{N1<tempo>}] [en + X_{N1<tempo>}]]$   $[[de + X_{N1<luogo>}] [en + X_{N1<luogo>}]]$
- Italiano:  $[[di + X_{N1<tempo>}] [in + X_{N1<tempo>}]]$   $[[di + X_{N1<luogo>}] [in + X_{N1<luogo>}]]$
- Spagnolo  $[[de + X_{N1<tempo>}] [en + X_{N1<tempo>}]]$   $[[de + X_{N1<luogo>}] [en + X_{N1<luogo>}]]$

Nelle tabelle che seguono si riportano i dati dell'italiano (tabella 16) e del francese (tabella 17), per il confronto del tipo di esemplari:

Semantica i = 'progressione temporale'				Semantica ii = 'progressione spaziale'			
	Sequenza (lemma)	Fq	Frequenza Normalizzata		Sequenza (lemma)	Fq	Frequenza Normalizzata
1.	<i>di volta in volta</i>	63326	1.079.820	1.	<i>di casa in casa</i>	2050	0.34956
2.	<i>di anno in anno</i>	16282	277.637	2.	<i>di città in città</i>	927	0.15807
3.	<i>di giorno in giorno</i>	10132	172.768	3.	<i>di paese in paese</i>	732	0.12482
4.	<i>di mese in mese</i>	1649	0.28118	4.	<i>di luogo in luogo</i>	364	0.06207
5.	<i>di momento in momento</i>	536	0.09140	5.	<i>di zona in zona</i>	132	0.02251

Tabella 16. Primi 5 esempi nel corpus itTenTen16

Semantica i = 'progressione temporale'				Semantica ii = 'progressione spaziale'			
	Sequenza (lemma)	Fq	Frequenza Normalizzata		Sequenza (lemma)	Fq	Frequenza Normalizzata
1.	<i>de temps en temps</i> 'di tanto in tanto'	74085	1.082.223	1.	<i>de ville en ville</i> 'di città in città'	1644	0.24015
2.	<i>de jour en jour</i> 'di giorno in giorno'	11299	165.054	2.	<i>de maison en maison</i> 'di casa in casa'	890	0.13001
3.	<i>de année en année</i> 'di anno in anno'	11239	164.178	3.	<i>d'île en île</i> 'di isola in isola'	695	0.10152
4.	<i>de heure en heure</i> 'di ora in ora'	1048	0.15309	4.	<i>de pays en pays</i> 'di paese in paese'	482	0.07041
5.	<i>de moi en moi</i> 'di mese in mese'	1026	0.14988	5.	<i>de lieu en lieu</i> 'di luogo in luogo'	438	0.06398

Tabella 17. Primi 5 esempi nel corpus frTenTen16<sup>40</sup>

Vale la pena notare che, per tutte le lingue considerate, i range paradigmatici relativi alle due semantiche sono piuttosto ridotti. Tuttavia, i *fillers* relativi alla semantica spaziale si mostrano più numerosi rispetto a quelli dal valore temporale; questo è dovuto al fatto che la costruzione spaziale può includere al suo interno elementi lessicali che designano metonimicamente luoghi (per es. il lessema *porta* in *di porta in porta* indica metonimicamente un edificio). Inoltre, vale la pena notare che dal confronto interlinguistico emerge una caratteristica degna di nota: la costruzione con valore temporale mostra elevate frequenze d'uso sia in italiano sia in francese, superando nettamente il valore spaziale in termini di occorrenze nel corpus. Questo può essere un indice di avvenuta lessicalizzazione e cristallizzazione delle sequenze temporali nelle due lingue. Tale ipotesi non può tuttavia applicarsi al caso dello spagnolo, dove i due valori si equivalgono in termini di frequenza.

## 5. Conclusioni

I meccanismi di combinazione lessicale hanno come esito insiemi di strutture largamente eterogenei. Alcune, tra di esse, mantengono caratteristiche simili alle configurazioni sintattiche, mentre altre si avvicinano per funzione alle unità lessicali monorematiche. I criteri tradizionalmente utilizzati in letteratura per la descrizione dei diversi fenomeni combinatori si mostrano non completamente sufficienti a descrivere l'importante variabilità di tali sequenze. I diversi parametri di classificazione tipicamente usati (per es. lessicali, semantici, sintattici, morfologici) vanno anzitutto descritti come proprietà graduali, che prevedano degli stadi intermedi, i cui tratti a volte non sono nettamente definibili. Inoltre, ai parametri tradizionali vanno necessariamente affiancati altri criteri, quali ad esempio la produttività, la predicibilità semantica, il livello di lessicalizzazione e l'apertura alla variazione paradigmatica. L'applicabilità di tali criteri è resa possibile attraverso l'esame di dati reali, e quindi attraverso una metodologia di analisi che consideri l'uso e la frequenza di occorrenza

<sup>40</sup> Si riportano i primi cinque esempi rappresentati nella tabella 9.



dei fenomeni. Infine, secondo queste generalizzazioni, i concetti di “combinazione di parola” e di “costruzione” devono necessariamente essere riconsiderati tenendo presente la prospettiva interlinguistica.

Seguendo queste considerazioni teoriche e metodologiche, è possibile disporre le combinazioni di parole lungo un *continuum*, che veda ai poli estremi due tipi di configurazioni:

- le *costruzioni parzialmente riempite*, caratterizzate da
  - i) variazione lessicale: presenza di una o più posizioni sintattiche vuote, saturabili attraverso materiale lessicale omogeneo,
  - ii) diversi gradi di idiomaticità: un significato non sempre compositivo, ma spesso costruzionale (vale a dire non trasparente, ma rintracciabile nella combinazione),
  - iii) una bassa coesione sintattica dei costituenti;
- le *costruzioni completamente riempite*, caratterizzate da
  - i) invariabilità lessicale,
  - ii) forte coesione sintattica,
  - iii) gradi di idiomaticità variabili.

Le prime sono sequenze sintagmatiche o frasali non pienamente lessicalizzate, dalle quali possono essere ricavati degli schemi più astratti, vere e proprie matrici (*templates*) attraverso le quali possono essere create nuove forme combinatorie dalle proprietà sintattico/semantiche simili e altamente predicibili sulla base della configurazione di partenza. Le seconde, all'estremo destro del *continuum*, sono sequenze pienamente lessicalizzate, che possono derivare da strutture produttive ma non ne realizzano, e non sono predicibili dal punto di vista semantico (la configurazione sintattica non veicola alcun tipo di informazione semantica e le sequenze sono spesso portatrici di un significato idiomatizzato e - apparentemente - idiosincratizzato).

Si rappresenta quanto detto nel grafico che segue, che vede nel polo sinistro del *continuum* le combinazioni di parole parzialmente riempite, e nel polo opposto le combinazioni completamente riempite.

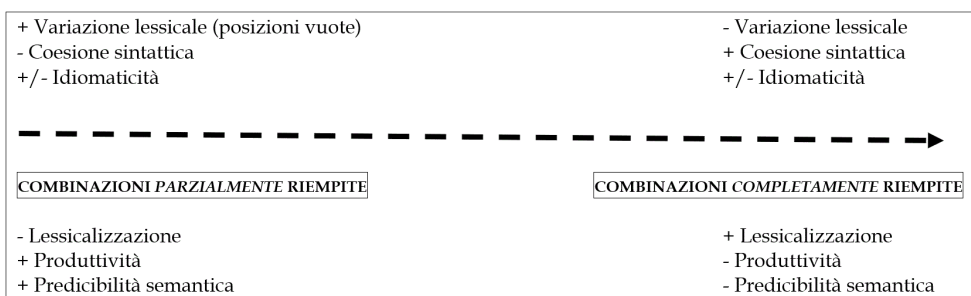


Figura 3. Continuum di lessicalizzazione

Nonostante presentino delle posizioni vuote, le CPR sono a tutti gli effetti delle strutture multiparola. Hanno pertanto lo stesso statuto di altri fenomeni combinatori, come le combinazioni completamente riempite. In diacronia, alcune di esse

possono cristallizzarsi e, attraverso un lento processo di lessicalizzazione, acquisire maggiore idiomatichità e fissità sintattica (il cosiddetto *figement* semantico e sintattico). La schematicità di alcune costruzioni può quindi diminuire col tempo, facendo sì che la costruzione non sia più rintracciabile. Altre costruzioni, invece, possono acquisire una maggiore schematicità e dar vita a configurazioni sintattico/semantiche dall'alta produttività.<sup>41</sup>

Anche se possono essere specifiche di una lingua, le CPR sono un fenomeno molto diffuso nelle lingue. E poiché difficili da rintracciare persino attraverso le metodologie di ricerca più recenti, tali tipi combinatori rappresentano un problema per l'ambito computazionale. Costituiscono anche un fenomeno poco rappresentato dal punto di vista lessicografico: sino ad ora sono state incluse solo in un'opera lessicografica dell'italiano, il dizionario combinatorio *CombiNet*<sup>42</sup>, e, sebbene ristretta, una loro rappresentazione si può osservare anche nel dizionario combinatorio dello spagnolo *REDES* (Bosque 2004).

## Abbreviazioni

AGG	aggettivo
N	nome
PART. PASS.	participio passato
PART. PRES.	participio presente
V	verbo

## Ringraziamenti

Sono grata a Vittorio Ganfi e ai revisori anonimi della rivista per le utili osservazioni alle versioni precedenti di questo contributo. Ringrazio Carmen Mellado Blanco, Raffaele Simone, Lunella Mereu, Anna Pompei e Francesca Di Salvo per i preziosi suggerimenti e le discussioni durante la conferenza EUROPHRAS 2019 – *Productive Patterns in Phraseology* e il ciclo di seminari *Linguaggio a Roma Tre*. Eventuali errori o imprecisioni sono da attribuire unicamente a me.

### Riferimenti bibliografici

- BALLY, Charles (1951[1909]), *Traité de stylistique française*, Paris: Klincksieck.
- BAREDDAL, Johanna (2008), *Productivity. Evidence from Case and Argument Structure in Icelandic*, Amsterdam-Philadelphia: Benjamins.
- BARONI, Marco *et al.* (2004), «Introducing the La Repubblica corpus: a large, annotated, TEI(XML)-compliant corpus of newspaper Italian», in LINO, M. T. *et al.* (a cura di), *Proceedings of the Fourth International Conference on Language Resources and Evaluation*, Paris: ELRA, 1771-1774.

<sup>41</sup> Cfr., ad esempio, Hoffmann – Trousdale (2011) per l'inglese, Piuanno – Ganfi (in stampa) per l'italiano, Di Salvo (2019) per il latino.

<sup>42</sup> Il dizionario *CombiNet* nasce dal Progetto PRIN 2010–2011 (n. 20105B3HE8): “CombiNet – Word Combinations in Italian: theoretical and descriptive analysis, computational models, lexicographic layout and creation of a dictionary”, finanziato dal Ministero dell'Istruzione e della Ricerca, a cui hanno partecipato l'Università Roma Tre, l'Università di Pisa e l'Università di Bologna. Per approfondimenti si rimanda a Simone – Piuanno (2017), Lenci *et al.* (2017), Castagnoli *et al.* (2015), Piuanno (2016b).

- BOAS, Hans C. (2010a), «The syntax-lexicon *continuum* in Construction Grammar. A case study of English communication verbs», *Belgian Journal of Linguistics* 24, 54-82.
- BOAS, Hans C. (2010b), *Contrastive Studies in Construction Grammar*, Amsterdam-Philadelphia, Benjamins.
- BOSQUE, Ignacio (2004), *REDES. Diccionario combinatorio del español contemporáneo. Las palabras en su contexto*, Madrid: Ediciones SM.
- BYBEE, Joan (2006), «From Usage to Grammar: The Mind's Response to Repetition», *Language* 82/4, 711-733.
- BYBEE, Joan (2010), *Language, Usage and Cognition*, Cambridge: Cambridge University Press.
- CASTAGNOLI, Sara *et al.*, «Towards a corpus-based online dictionary of Italian Word Combinations Automatic Knowledge Acquisition for Lexicography», *COST ENeL WG3 meeting, Herstmonceux Castle, 13 August 2015* [consultabile all'indirizzo <[https://www.elexicography.eu/wp-content/uploads/2015/10/Paper\\_Castagnoli\\_et\\_al\\_ENeL\\_final.pdf](https://www.elexicography.eu/wp-content/uploads/2015/10/Paper_Castagnoli_et_al_ENeL_final.pdf)> (13/6/2020).
- CASTAGNOLI, Sara *et al.* (2016), «POS-patterns or Syntax? Comparing methods for extracting Word Combinations», in CORPAS PASTOR, G. (a cura di), *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)*, Geneva: Tradulex, 101-114.
- DI SALVO, Francesca (2019), «CVS in latino: (*in*) numero / (*in*) loco habeo», in POMPEI, A. – MEREU, L. (a cura di), *Verbi Supporto: Fenomeni e Teorie*, München: Lincom Europa, 155-185.
- DOBROVOL'SKIJ, Dmitrij (2011), «Phraseologie und Konstruktionsgrammatik», in LASCH, A. – ZIEM; A. (a cura di). *Konstruktionsgrammatik III. Aktuelle Fragen und Lösungsansätze*, Tübingen: Stauffenburg, 111-130.
- FILLMORE, Charles J. – KAY, Paul – O'CONNOR, Mary Catherin (1988), «Regularity and idiomatity in grammatical constructions: the case of let alone», *Language* 64, 501-538.
- FREI, Henri (1969[1929]), *La Grammaire des fautes*, Paris-Genève-Leipzig: Geuthner-Kundig-Harrassowitz.
- GANEI, Vittorio – PIUNNO, Valentina (2017), «Preposizioni complesse in italiano antico e contemporaneo. Grammaticalizzazione, schematismo e produttività», *Archivio Glottologico Italiano* 102/2, 184-204.
- GOLDBERG, Adele (1995), *Constructions. A construction Grammar Approach to Argument Structures*, Chicago, The University of Chicago Press.
- GOLDBERG, Adele (2003), «Constructions: a new theoretical approach to language», *TRENDS in Cognitive Sciences* 7/5, 219-224.
- GOLDBERG, Adele (2006), *Constructions at work*, Oxford: Oxford University Press.
- GROSS, Gaston (1996), *Les expressions figées en français*, Paris: Ophirys.
- HOFFMANN, Thomas – TROUSDALE, Graeme (2011), «Variation, change and constructions in English», *Cognitive Linguistics* 22/1, 1-23.
- LAMIROY, Béatrice (2003), «Les notions linguistiques de figement et de contrainte», *Linguisticae Investigationes* 26/1, 1-14.
- LENCI, Alessandro *et al.* (2017), «How to harvest Word Combinations from corpora. Methods, evaluation and perspectives», *Studi e Saggi Linguistici* 55/2, 45-68.
- LYDING, Verena *et al.* (2014), «The PAISÀ Corpus of Italian Web Texts», in BILDHAUER, F. – SCHÄFER, R. (a cura di), *Proceedings of the 9th Web as Corpus Workshop (WaC-9)*, Gothenburg: Association for Computational Linguistics, 36-43.

- MASINI, Francesca (2013), *Parole sintagmatiche in italiano*, Cesena – Roma: Caissa Italia.
- MEJRI, Salah (1997), *Le figement lexical. Descriptions linguistiques et structuration sémantique*, Tunis: Publications de la Faculté des lettres de la Manouba.
- MELLADO BLANCO, CARMEN (in stampa), «Esquemas fraseológicos y construcciones fraseológicas en el continuum léxico-gramática», in SINNER, C. – TABARES PLASENCIA, E. – MONTORO DEL ARCO, E. (a cura di), *Clases y categorías en la fraseología española*, Frankfurt am Main: Peter Lang.
- MELLADO BLANCO, Carmen – LÓPEZ MEIRAMA, Belén (2017), «El fraseologismo por momentos: principales valores semánticos y algunos apuntes diatópicos», *RILCE* 33/2, 648-670.
- MELLADO BLANCO, Carmen – LÓPEZ MEIRAMA, Belén (2018), «Las construcciones [de X a Y] y [de X a X] realizaciones idiomáticas y no tan idiomáticas», in DÍAZ FERRO, M. *et al.* (a cura di), *Actas do XIII Congreso Internacional de Lingüística Xeral*, Vigo: Universidad de Vigo, 576-583, <<https://dialnet.unirioja.es/servlet/articulo?codigo=6506540>> (9/6/2020).
- MICHAELIS, Laura A. (in stampa, a), «Meanings of constructions», *Oxford Research Encyclopedia of Linguistics*.
- MICHAELIS, Laura A. (in stampa, b), «Constructions are Patterns and so are Fixed Expressions», in BUSSE, B. – MÖHLIG-FALKE, R. (a cura di), *Patterns in Language and Linguistics*, Berlin: Mouton de Gruyter.
- PASSARO, Lucia C. – LENCI, Alessandro (2016), «Extracting Terms with EXTra», relazione presentata al convegno *EUROPHRAS 2015 – Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives*, Malaga, 29 giugno -1 luglio 2015.
- PIUNNO, Valentina (2015), «Sintagmi Preposizionali come Costruzioni Aggettivali», *Studi e Saggi Linguistici* 53/1, 65-98.
- PIUNNO, Valentina (2016a), «Combinazioni di parole parzialmente riempite. Formati e rappresentazione lessicografica», *Studia de Cultura* 9/3, 103-113.
- PIUNNO, Valentina (2016b), «Multiword Modifiers in Romance languages. Semantic formats and syntactic templates», *Yearbook of Phraseology* 7, 3-34.
- PIUNNO, Valentina (2016c), «CombiNet. A Corpus-based Online Database of Italian Word Combinations», in KERNERMAN, I. (a cura di), *GLOBALEX 2016. Lexicographic resources for human language technology*, Paris: ELRA, 45-51.
- PIUNNO, Valentina (2018), *Sintagmi preposizionali con funzione aggettivale e avverbiale*, München: LINCOM.
- PIUNNO, Valentina – GANFI, Vittorio (in stampa), «Synchronic and diachronic analysis of prepositional multiword modifiers across Romance languages», in GREZKA, A. – MOGORRÓN, P. (a cura di), *Special Issue of Linguisticae Investigationes. Les variations diatopiques dans les expressions figées*.
- PIUNNO, Valentina – GANFI, Vittorio (2019), «Usage-based account of Italian Complex Prepositions denoting the Agent», *Revue Romane* 54/1, 141-175.
- POTTIER, Bernard (1974[1967]), *Présentation de la linguistique. Fondements d'une théorie*, Paris: Klincksieck.
- POTTIER, Bernard (1985), *Linguistique générale. Théorie et description*, Paris: Klincksieck.
- SECHEHAYE A. 1950 [1926], «Essai sur la structure logique de la phrase», *Collection linguistique* 20, Paris: Champion.

- SIMONE, Raffaele (2003), «Masdar, 'ismu-al-marrati et la frontière verbe/nom», in GIRON, J. L. – HERRERO, F. J. – IGLESIAS, S. (a cura di), *Estudios ofrecidos al profesor José Jesús de in stampa Bustos Tovar*, Madrid: Editorial Complutense, vol. 1, 901-918.
- SIMONE, Raffaele (2006), «Nominales sintagmáticos y no-sintagmáticos», in DE MIGUEL APARICIO, E. – AZUCENA PALACIOS, A. – SERRADILLA CASTAÑO, A. (a cura di), *Estructuras léxicas y estructuras del léxico*, Berlin: Peter Lang, 221-241.
- SIMONE, Raffaele (2007), «Constructions and categories in verbal and signed languages», in PIZZUTO, E. – PIETRANDREA, P. – SIMONE, R. (a cura di), *Verbal and Signed Languages. Comparing Structures, Constructs, and Methodologies*, Berlin-New York: Mouton de Gruyter, 198-252.
- SIMONE Raffaele (2008), «Verbi sintagmatici come costruzione e come categoria», in CINI, M. (a cura di), *I verbi sintagmatici in italiano e nelle varietà dialettali. Stato dell'arte e prospettive di ricerca*, Berlin: Peter Lang, 13-30.
- SIMONE, Raffaele (2017), Word as a stratification of formats, in D'Alessandro, R. et al. (a cura di), *Di tutti i colori. Studi linguistici per Maria Grossmann*, Utrecht: Utrecht University Repository, 335-347.
- SIMONE, Raffaele – MASINI, Francesca (2007), «Support nouns and verbal features: a case study from Italian», in GREZKA, A. – MARTIN-BERTHET, F. (a cura di), *Verbes et classes sémantiques*, *Verbum* 29, 143-172.
- SIMONE, Raffaele – MASINI, Francesca. (2014), «On Light Nouns», in SIMONE, R. – MASINI, F. (a cura di), *Word classes: nature, typology and representations*, Amsterdam-Philadelphia: Benjamins, 51-73.
- SIMONE, Raffaele – PIUNNO, Valentina (2017a), «Introduction», *Studi e Saggi Linguistici. Special Issue Word Combinations: phenomena, methods of extraction, tools* 55/2, 9-11.
- SIMONE, Raffaele – PIUNNO, Valentina (2017b), «Combinazioni di parole che costituiscono entrata. Rappresentazione lessicografica e aspetti lessicologici», *Studi e Saggi Linguistici* 55/2, 13-44.
- STEYER, Kathrin (2018, a cura di), *Sprachliche Verfestigung. Wortverbindungen, Muster, Phrasem-Konstruktionen*, Tübingen: Narr.
- SVENSSON, Maria Helena (2004), *Critères de figement*, Umeå, Umeå University.
- VOGHERA, Miriam (1994), «Lessemi complessi: percorsi di lessicalizzazione a confronto», *Lingua e Stile* 29/2, 185-214.
- VOGHERA, Miriam (2004), «Polirematiche», in GROSSMANN, M. – RAINER, F. (a cura di), *La formazione delle parole in italiano*, Tübingen: Niemeyer, 56-69.
- ZIEM, Alexander (2008), *Frames und sprachliches Wissen. Kognitive Aspekte der semantischen Kompetenz*, Berlin-New York: Mouton de Gruyter.
- ZIEM, Alexander (2018), «Construction Grammar meets Phraseology: eine Standortbestimmung», *Linguistik Online*, 90(3), 3-18.
- ZIEM, Alexander – LASCH, Alexander (2013), *Konstruktionsgrammatik: Konzepte und Grundlagen gebrauchsbasierter Ansätze*, Berlin-New York: Mouton de Gruyter.

### Risorse sul web

- Corpus del Español Actual*, <<http://spanishfn.org/tools/cea/english>> [13/6/2020].
- Corpus La Repubblica*, <[https://corpora.dipintra.it/public/run.cgi/corp\\_info?corp\\_name=repubblica](https://corpora.dipintra.it/public/run.cgi/corp_info?corp_name=repubblica)> [13/6/2020].
- Corpus Paisà*, <<https://www.corpusitaliano.it>> [13/6/2020].
- SketchEngine*, <<http://www.sketchengine.eu>> [13/6/2020].

