Polirematiche ‘di colore’ in italiano: uno studio quantitativo

Francesca Masini

Abstract

This paper investigates the presence and use of basic color terms within Italian multiword expressions, in particular phrasal lexemes (e.g. *libro bianco* ‘white book’, *vedere nero* ‘be pessimistic’). The starting hypothesis is that most frequent color terms, especially those that are ranked higher in Berlin & Kay’s (1969) implicational hierarchy, are more widely used within multiword units. In order to test the hypothesis, data from the largest dictionary available for the Italian language are gathered, together with semi-automatically collected data from corpora, which considerably enrich the sample. The hypothesis is confirmed: more frequent, ‘higher-ranked’ color terms (viz. *bianco* ‘white’, *nero* ‘black’, *rosso* ‘red’) are found in a larger set of phrasal lexemes as compared to less frequent color terms.

KEYWORDS: color terms • frequency • implicational hierarchy • Italian • multiword expressions • phrasal lexemes

1. Introduzione

In questo articolo esporrò alcuni dati quantitativi relativi alla presenza dei termini di colore nelle polirematiche italiane, nella speranza che la scelta dell’argomento non risulti troppo ardita, ma anzi gradita alla festeggiata, Maria Grossmann, vista la sua nota predilezione per gli studi sul lessico dei colori (cfr., su tutti, Grossmann 1988).

Con il termine ‘polirematiche’ (De Mauro 1999, Voghera 2004, Masini 2011) mi riferisco a elementi lessicali complessi, formati da più di una parola e appartenenti a diverse parti del discorso, che presentano un certo grado di fissità interna (sintagmatica e paradigmatica) e un significato unitario e convenzionalizzato (es. *bandiera bianca*, *libro nero*, *filo rosso*, *casco blu*, *telefono azzurro*, *carta verde*, *pagine gialle*, *materia grigia*, *maglia rosa*). Verranno prese in considerazione anche espressioni che possono essere (o sono state) considerate composti (es. *Croce rossa*, *verde bottiglia*, *blu elettrico*).

Lo scopo di questo studio è mostrare come i termini di colore siano impiegati nelle parole polirematiche della lingua italiana. L’ipotesi di partenza è che i termini più frequenti siano quelli più utilizzati all’interno di questi lessemi complessi. Secondo Voghera (2004: 66), infatti, l’alta frequenza di un lessema favorirebbe la sua occorrenza all’interno di formazioni polirematiche, perché i lessemi molto frequenti sono solitamente polisemici e più vaghi di quelli meno frequenti, e possono quindi comparire in più contesti. Sebbene Voghera qui parli di verbi, il ragionamento può essere esteso anche ad altre parti del discorso. Allo stesso tempo, ci aspettiamo che i termini più usati all’interno delle polirematiche siano quelli che si trovano più in alto (o a sinistra) nella nota scala implicazionale di Berlin & Kay (1969), che contiene gli 11 termini di colore considerati ‘di base’ a livello interlinguistico.

L’articolo è strutturato come segue. Nel § 2 descriverò la metodologia e le risorse utilizzate nello studio. Il § 3 illustrerà invece i risultati dell’analisi. Nel § 4 trarrò alcune conclusioni.

2. Metodologia

La base di dati di partenza per questo studio è costituita dall’elenco dei 12 termini di colore in (1), i quali corrispondono agli 11 termini basici della scala implicazionale di Berlin & Kay (1969: 4), riportata in (2). L’unità in più (12 vs. 11) è dovuta al fatto che *blue* può corrispondere, in italiano, a due termini di colore entrambi considerati basici, ovvero *blu* e *azzurro* (Grossmann & D’Achille 2016).

- (1) bianco, nero, rosso, verde, giallo, blu, azzurro, marrone, viola, rosa, arancione, grigio

(2) white | black > red > green | yellow > blue > brown > purple | pink | orange | grey

L'uso di questi termini nelle polirematiche italiane è stato investigato utilizzando le seguenti risorse:

- il *Nuovo De Mauro* (d'ora in poi NDM), che, oltre ad essere liberamente disponibile online¹, lemmatizza le polirematiche;
- il corpus di italiano giornalistico *la Repubblica* (d'ora in poi laR), che contiene 16 annate dell'omonimo quotidiano per un totale di (circa) 380 milioni di token (Baroni et al. 2004)²;
- i dati, estratti automaticamente dal corpus laR, elaborati all'interno del progetto *CombiNet* sulle combinazioni di parole in italiano.³

Innanzitutto, ho estratto dal NDM tutte le polirematiche relative ai 12 lemmi elencati in (1) (cfr. § 3.1), tenendo traccia della loro marca d'uso, in particolare della marca TS, che identifica i termini “legati a un uso marcatamente o esclusivamente tecnico-specialistico”, noti “soprattutto in rapporto a particolari attività, tecnologie, scienze” (De Mauro 2005: 60). Trattandosi di una risorsa internet, preciso che l'estrazione dei dati è avvenuta nel corso del mese di luglio 2016. Successivamente, ho calcolato la frequenza complessiva degli stessi 12 lemmi nel corpus laR (cfr. § 3.2), per poi passare ai dati *CombiNet* (cfr. § 3.3). Questi ultimi consistono in liste di combinazioni di parole estratte automaticamente da corpora, utilizzando tecniche diverse, e ordinate tramite misure di associazione statistica (ma è disponibile anche il dato di frequenza). In particolare, i dati su cui mi sono basata per questo studio sono stati estratti dal corpus laR tramite una tecnica basata su sequenze lineari di parti del discorso (POS). Le sequenze POS utilizzate all'interno del progetto *CombiNet* sono 122 e corrispondono a combinazioni, fino a 5 elementi, di struttura diversa. Si vedano, ad esempio, le sequenze POS illustrate in (3).⁴

- (3) a. N + PREP + N + AGG (→ *contratto a tempo determinato*)
b. V + ARTDET + N (→ *alzare il gomito*)
c. AGG + PREP + N (→ *nuovo di zecca*)
d. PREP + N + AGG (→ *di umore nero*)

Tali sequenze sono state immesse nel software EXTra (Passaro & Lenci 2016), che ha estratto tutte le combinazioni corrispondenti ai 122 pattern con frequenza <5, ordinandole poi per associazione statistica; la misura di associazione statistica utilizzata è la *Log Likelihood* (il calcolo è effettuato sui lemmi, non sulle forme). L'estrazione è stata effettuata su 4 tornate, suddividendo le sequenze POS in 4 gruppi, ovvero sequenze POS che identificano: i) sintagmi nominali e aggettivali; ii) sintagmi verbali; iii) sintagmi preposizionali; iv) e infine sequenze miste/residuali.⁵

Le liste di stringhe estratte da EXTra sono ovviamente liste di ‘candidati’, in cui troveremo sia combinazioni di parole che corrispondono a polirematiche (o a collocazioni), sia combinazioni semplicemente molto frequenti che però non costituiscono espressioni fisse. I dati così estratti, quindi, richiedono uno spoglio manuale.

Per questo specifico studio, ho estratto, dalla base di dati EXTra appena descritta, tutte le stringhe contenenti uno qualsiasi dei 12 lemmi di colore elencati in (1). I dati così ottenuti sono stati incrociati con quelli del NDM, per verificare quali polirematiche registrate nel NDM siano ricavabili tramite

¹ URL: <http://dizionario.internazionale.it>.

² Lo studio si basa su una versione del corpus *la Repubblica* etichettata per parti del discorso tramite la procedura descritta in Dell'Orletta (2009) e annotata sintatticamente tramite DeSR (Attardi & Dell'Orletta 2009). Si tratta dunque di una versione non identica a quella disponibile all'indirizzo <http://sslimit.unibo.it/repubblica>.

³ Progetto PRIN 2010-2011 n. 20105B3HE8 (*Combinazioni di parole in italiano: analisi teorica e descrittiva, modelli computazionali, studio di un layout lessicografico dedicato e realizzazione di un dizionario di combinazioni*) finanziato dal Ministero dell'Istruzione, dell'Università e della Ricerca e diretto da Raffaele Simone e Alessandro Lenci. URL: <http://combinet.humnet.unipi.it>.

⁴ Lista delle abbreviazioni: AGG = aggettivo; ARTDET = articolo determinativo; AVV = avverbio; N = nome; PREP = preposizione; V = verbo.

⁵ Per maggiori dettagli sulle tecniche di estrazione dei dati all'interno di *CombiNet* si vedano i contributi di Castagnoli et al. (2016) e Lenci et al. (2017).

tecniche di estrazione semi-automatica. Allo stesso tempo, ho effettuato un primo, preliminare spoglio manuale delle liste ottenute da EXTra, allo scopo di verificare se siano presenti combinazioni non registrate nel NDM (cfr. § 3.3).

3. Dati

3.1 Le polirematiche ‘di colore’ nel *Nuovo De Mauro*

Come già accennato, nel NDM online le polirematiche sono lemmatizzate (più precisamente, sono registrate come sottolemmi dei lemmi monorematici in esse contenuti), e quindi facilmente recuperabili. In coda alla voce per il lemma *verde*, quindi, troveremo una sezione ‘polirematiche’ in cui figurano tutte le polirematiche in cui compare *verde*, ad esempio *bollino verde*, *al verde*, *verde bottiglia*. Quest’ultima espressione sarà a sua volta contenuta, come sottolemma polirematico, anche nell’entrata *bottiglia*. I dati ottenuti dallo spoglio del NDM sono riassunti nella Tabella 1 (l’elenco complessivo è riportato in Appendice).

TABELLA 1. Polirematiche contenenti termini di colore nel *Nuovo De Mauro*.

COLORE	N° POLIREMATICHE NEL NDM	N° POLIREMATICHE MARCATE COME TS	POSIZIONE RISPETTO A BERLIN/KAY
bianco	131	67 (51,14%)	=
nero	91	41 (45,05%)	=
rosso	75	52 (69,33%)	=
verde	44	17 (38,63%)	=
giallo	29	16 (55,17%)	=
azzurro	19	13 (68,42%)	=
grigio	18	13 (72,22%)	↑
blu	14	5 (35,71%)	↓
rosa	11	4 (36,36%)	↑
marrone	0	-	↓
arancione	0	-	=
viola	0	-	=

Come possiamo vedere, la prima colonna contiene i termini di colore, ordinati per numero di polirematiche registrate nel NDM (colonna 2); la terza colonna mostra invece il numero assoluto e la percentuale (piuttosto elevata) di polirematiche marcate come TS. Come si può notare, i colori che presentano il numero maggiore di polirematiche sono *bianco*, *nero* e *rosso*, seguiti da *verde* e *giallo* e poi da *azzurro*, *grigio*, *blu* e *rosa*. Non ci sono polirematiche relative a *marrone*, *arancione* e *viola*. Come evidenziato nell’ultima colonna a destra, quest’ordine è pressoché corrispondente alla scala implicazionale di Berlin & Kay (1969), salvo la parte medio-bassa, in cui notiamo qualche avvicendamento (es. *grigio* è più ‘alto’ di *blu*, *rosa* è più ‘alto’ di *marrone*). Per quanto riguarda la posizione relativa di *grigio* e *blu*, va ricordato che *blu* è solo uno dei due termini che si riferiscono a *blue* in italiano: se lo sommiamo ad *azzurro*, il numero totale supera quello di *grigio* (ma anche quello di *giallo*).

Le polirematiche registrate nel NDM sono quasi tutte di natura nominale, come mostrato nella Tabella 2 (che contiene un esempio per ogni tipo). Tutte le altre parti del discorso sono decisamente meno numerose. Vale la pena notare che, tra queste ultime, quella più rappresentata è, forse inaspettatamente, quella dei verbi. Inoltre, man mano che scendiamo lungo la scala dei termini di colore, troviamo sempre meno varietà a livello di categorie in uscita rappresentate, un altro possibile indice della maggiore combinabilità dei termini più frequenti e più basilari.

3.2 Frequenza dei termini di colore nel corpus *la Repubblica*

Per avere un’indicazione di massima dell’incidenza complessiva dei 12 termini di colore oggetto della presente analisi sul lessico dell’italiano (indipendentemente dalla loro occorrenza nelle polirematiche), ho estratto le frequenze complessive di tali termini all’interno del corpus laR. Il computo è stato effettuato per lemma e senza restrizioni in termini di POS (nome vs. aggettivo); preciso che si tratta di dati

grezzi, non sottoposti a controllo manuale, pertanto è possibile che si presentino delle anomalie (cfr. *infra*). I dati sono riportati nella Tabella 3.

Anche in questo caso, notiamo come l'ordine dei termini di colore per frequenza ricalchi grosso modo la scala implicazionale di Berlin & Kay, specialmente nella parte alta. Nella parte centrale osserviamo qualche cambio di posizione, che però potrebbe essere riconducibile alla natura, poco controllata, dei dati. La posizione 'alta' di *azzurro*, per esempio, potrebbe essere dovuta all'alta frequenza (a maggior ragione in un corpus giornalistico) della forma *azzurri* nell'accezione di 'italiani' (es. *atleti azzurri, gli azzurri*, ecc.).

TABELLA 2. Polirematiche 'di colore' e classi di parole nel Nuovo De Mauro.

COLORE	N	V	N AGG	AGG	AGG AVV	AVV	TOTALI
bianco	116 <i>sciopero bianco</i>	7 <i>firmare in bianco</i>	1 <i>russo bianco</i>	3 <i>in bianco e nero</i>	2 <i>in bianco</i>	2 <i>a pane bianco</i>	131
nero	83 <i>bucio nero</i>	4 <i>fare nero</i>	1 <i>nero fumo</i>	2 <i>di umore nero</i>	1 <i>in nero</i>	-	91
rosso	70 <i>brigiate rosse</i>	1 <i>vedere rosso</i>	1 <i>pelle rossa</i>	2 <i>a luci rosse</i>	1 <i>in rosso</i>	-	75
verde	37 <i>onda verde</i>	1 <i>vedere i sorci verdi</i>	4 <i>verde acido</i>	-	1 <i>al verde</i>	1 <i>in verde</i>	44
giallo	28 <i>cartellino giallo</i>	-	-	-	-	1 <i>in guanti gialli</i>	29
azzurro	19 <i>pesce azzurro</i>	-	-	-	-	-	19
grigio	18 <i>materia grigia</i>	-	-	-	-	-	18
blu	13 <i>casco blu</i>	-	1 <i>blu elettrico</i>	-	-	-	14
rosa	10 <i>salsa rosa</i>	1 <i>vedere rosa</i>	-	-	-	-	11
TOTALE	394	14	8	7	5	4	432

TABELLA 3. Dati di frequenza (grezzi) per i 12 lemmi di colore nel corpus IaR.

COLORE	FREQUENZA TOTALE NEL CORPUS IaR	POSIZIONE RISPETTO A BERLIN/KAY
nero	72358	=
bianco	52163	=
rosso	45122	=
verde	34982	=
azzurro	26198	↑
giallo	21960	↓
blu	15580	=
rosa	14407	↑
grigio	12184	↑
viola	5486	↑
marrone	1280	↓
arancione	1106	=

3.3 I dati *CombiNet*

In quest'ultimo paragrafo discuteremo i dati *CombiNet*, estratti automaticamente tramite EXTra (cfr. § 2). Uno spoglio completo dei dati EXTra va al di là dello scopo di questo studio: mi limiterò qui ad illu-

strare alcuni dati e fatti significativi, concentrandomi sui risultati ordinati più in alto nelle liste generate da EXTra (cfr. *infra*).

Come possiamo notare dalla Tabella 4, le stringhe totali⁶ estratte da EXTra variano notevolmente a seconda del colore: abbiamo ben 1301 combinazioni per *nero*, ma solo 13 per *arancione*. Ricordiamo che si tratta di ‘candidati’, di stringhe considerate statisticamente significative: non tutte corrisponderanno necessariamente a polirematiche.

TABELLA 4. Dati estratti da EXTra a confronto con il *Nuovo De Mauro*.

COLORE	N° STRINGHE ESTRATTE DA EXTRA	N° POLIREMATICHE NEL NDM	N° POLIREMATICHE DEL NDM ESTRATTE ANCHE DA EXTRA	N° POLIREMATICHE DEL NDM NON ESTRATTE DA EXTRA	
				TOTALE	MARCATO TS
nero	1301	91	62 (68,13%)	29 (31,87%)	24 (82,75%)
bianco	1086	131	60 (45,80%)	71 (54,20%)	53 (74,64%)
rosso	835	75	27 (36,00%)	48 (64,00%)	40 (83,33%)
azzurro	365	19	6 (31,58%)	13 (68,42%)	10 (76,92%)
verde	297	44	17 (38,64%)	27 (61,36%)	15 (55,55%)
blu	284	14	9 (64,29%)	5 (35,72%)	4 (80,00%)
giallo	230	29	12 (41,38%)	17 (58,62%)	9 (52,94%)
grigio	220	18	4 (22,22%)	14 (77,78%)	13 (92,85%)
viola	94	-	-	-	-
rosa	83	11	5 (45,45%)	6 (54,55%)	2 (33,33%)
arancione	13	-	-	-	-
marrone	0	-	-	-	-
TOTALE	4808	432	202 (46,76%)	230 (53,24%)	170 (73,91%)

Incrociando i dati di EXTra con la lista di polirematiche registrate nel NDM, osserviamo che quasi la metà delle polirematiche del NDM è presente anche nelle liste EXTra (202 su 432). Abbiamo invece 230 polirematiche (il 53,24% del totale) che non vengono estratte da EXTra. Va tuttavia sottolineato che la stragrande maggioranza di queste (73,91%) è marcata come TS nel NDM (es. *bianco di zinco* TS chim., *effluente nero* TS idraul., *trombo rosso* TS med.), come illustrato nell’ultima colonna a destra della Tabella 4. Trattandosi di termini usati in vari ambiti tecnico-specialistici, non stupisce che non ce ne sia traccia in un corpus di italiano giornalistico come laR, o che siano talmente poco frequenti da non rientrare nel processo di estrazione messo a punto per *CombiNet*.

Allo stesso tempo, le liste prodotte da EXTra contengono diverse polirematiche che non sono invece presenti nel NDM. Nei paragrafi che seguono illustrerò i risultati dell’analisi delle liste relative a ciascuno dei 12 termini di colore, mettendo in evidenza le polirematiche che si collocano più in alto, ovvero quelle caratterizzate da una maggiore associazione statistica, e quelle che non sono presenti nella nostra risorsa lessicografica di riferimento. Le polirematiche sono state individuate tramite spoglio manuale seguendo un criterio di coesione semantica. Non tutte le polirematiche sono caratterizzate da un significato completamente idiomatico – alcune sono decisamente opache (es. *pecora nera*, *telefono azzurro*), altre sono molto più trasparenti (es. *sposarsi in bianco*, *pepe verde*) – ma tutte hanno un significato *convenzionalizzato*, atto alla denominazione di un referente stabile.⁷ Nello spoglio mi sono limitata ai primi 100 risultati di EXTra per ciascuna delle 4 tornate di estrazione (cfr. nota 5, § 2), nei casi in cui, ovviamente, il numero di candidati estratti fosse >100.⁸ Nelle pagine che seguono chiamerò questo sottoinsieme di

⁶ Con ‘stringhe totali’ intendo l’unione delle 4 tornate di estrazione descritte nel § 2.

⁷ La convenzionalizzazione porta tipicamente con sé un surplus semantico. Un *panno verde*, per esempio, è effettivamente un panno di colore verde, ma non è un panno qualunque: è un panno di colore verde che si usa in un determinato contesto (quello dei giochi da tavolo). Un’espressione come *sposarsi in bianco* è facilmente interpretabile come ‘sposarsi con un abito bianco’, ma a ben vedere contiene molte più informazioni, come ad esempio il fatto che l’abito bianco in questione si riferisce unicamente alla sposa e che la cerimonia è di stampo tradizionale (il bianco simboleggerebbe la ‘purezza’ della sposa): se mi sposassi in comune dopo 15 anni di convivenza al cospetto dei soli testimoni e per l’occasione indossassi un tailleur bianco difficilmente direi che mi sono sposata in bianco.

⁸ In nessun caso raggiungiamo la teorica cifra di 400 candidati totali (100 per ogni tornata). Se la prima tornata – quella relativa alle sequenze POS che identificano sintagmi nominali e aggettivali – genera per tutti i colori molti risultati, su cui è stato

dati ‘campione’, per comodità. Fa eccezione, come vedremo, la ricerca di espressioni in cui il termine di colore funge da testa, che è stata effettuata sull’intero set di dati. Nel riportare gli esempi, seguirò l’ordine in cui compaiono nelle liste EXTra, partendo quindi dal più statisticamente significativo (che non necessariamente coincide con il più frequente).

3.3.1 *Nero*

All’interno delle liste EXTra per *nero* troviamo 62 delle 91 polirematiche registrate nel NDM. Circa un terzo di queste 62 si trova nel campione analizzato. Le espressioni in questione sono riportate in (4).

- (4) a. bianco e nero, scatola nera, bestia nera, fumata nera, camicia nera, maglia nera, messa nera, marea nera, borsa nera, cravatta nera, anima nera, pantera nera, bandiera nera, penna nera, continente nero, libro nero, basco nero, umorismo nero, pozzo nero, pecora nera (di famiglia), vedova nera, perla nera
- b. vedere nero
- c. in bianco e nero, in nero, di umore nero.

Nello stesso campione troviamo anche un discreto numero di polirematiche che non sono presenti nel NDM:

- (5) a. bianco o nero, venerdì nero, occhio nero, giornata nera, umore nero, toghe nere, muso nero, incazzato nero
- b. vestire di nero, lavorare in nero, pagare in nero, vedere un futuro nero, votare nero, chiudere in nero
- c. di pelle nera, in nero, (essere) in crisi nera.

Se consideriamo l’intero set di dati generato da EXTra, e non solo il campione qui preso in esame, troviamo anche polirematiche in cui il termine di colore costituisce la testa dell’espressione:

- (6) nero corvo, nero fumo, nero lucido, nero nero.

Le combinazioni del tipo *nero lucido* (Colore+AGG) sono normalmente considerate composti e “indicano una tonalità particolare del colore designato dalla testa” (D’Achille & Grossmann 2010: 409). Anche il tipo *nero fumo* (Colore+N) viene spesso analizzato come un composto, di tipo appositivo: qui la tonalità è specificata tramite una comparazione con il colore attribuito per eccellenza o al referente del nome o a un’entità ad esso strettamente legata (cfr. D’Achille & Grossmann 2013). Infine, nel caso di *nero nero* siamo chiaramente di fronte a una reduplicazione totale (Colore_i+Colore_i), che come ci ricordano D’Achille & Grossmann (2010: 407) può sia riferirsi al “punto considerato focale” se il colore in questione è usato in senso proprio (es. *labbra rosse rosse*), sia veicolare “intensificazione della qualità” se l’aggettivo di colore è usato in senso traslato (es. *partita grigia grigia*).

3.3.2 *Bianco*

Per *bianco* abbiamo ben 131 polirematiche registrate nel NDM, di cui 60 recuperate anche attraverso EXTra. Nel nostro campione ne troviamo quasi la metà:

- (7) a. bianco e nero, scheda bianca, carta bianca, cambiale in bianco, semestre bianco, fumata bianca, polvere bianca, (mettere) nero su bianco, vino bianco, mosca bianca, golpe bianco, squalo bianco, bianco e rosso, assegno in bianco, tartufo bianco, notte bianca, settimana bianca, voce bianca, camice bianco, bandiera bianca

necessario operare il taglio a quota 100, le stringhe relative ai sintagmi verbali e preposizionali sono meno numerose (in particolare quelle verbali), spesso <100. I risultati estratti per la quarta tornata (residuale) sono invece praticamente nulli.

- b. alzare bandiera bianca, firmare in bianco, (far) venire i capelli bianchi, andare in bianco
- c. in bianco e nero, all'arma bianca, di punto in bianco, in bianco.

A queste si aggiunge un altro discreto numero di polirematiche non presenti nel NDM, tra cui figurano anche alcune espressioni idiomatiche verbali e preposizionali:

- (8) a. bianco o nero, delega in bianco, balena bianca⁹, razza bianca, notte in bianco, cooperative bianche, dama bianca, polverina bianca, coop bianche
- b. avere i capelli bianchi, indossare il camice bianco, avere la pelle bianca, delegare in bianco, sposarsi in bianco
- c. di pelle bianca, di razza bianca.

Nel campione in esame troviamo solo un caso di combinazione che presenta *bianco* come testa: *bianco accecante*. Se però cerchiamo nel resto dei risultati, otteniamo un insieme più ampio di dati:

- (9) bianco accecante, bianco abbagliante, bianco virginale, bianco abbacinante, bianco immacolato, bianco avorio, bianco sporco, bianco latte, bianco candido.

3.3.3 Rosso

Delle 27 polirematiche registrate nel NDM e presenti anche nelle liste EXTra, 19 compaiono nel nostro campione (10).

- (10) a. cartellino rosso, filo rosso, bianco e rosso, brigate rosse, stella rossa, libretto rosso, camicia rossa, khmer rossi, armata rossa, linea rossa, guardia rossa, pesce rosso, disco rosso, carne rossa, basco rosso, abete rosso
- b. vedere rosso
- c. a luci rosse, in rosso.

La stringa *a luci rosse* viene recuperata da EXTra sia come stringa indipendente, sia come modificatore di diversi nominali (es. *film a luci rosse*, *quartiere a luci rosse*, *cinema a luci rosse*).

Sempre nel nostro campione di dati troviamo alcune combinazioni preferenziali, come ad esempio *mazzo di rose rosse*, nonché diverse polirematiche non registrate nel NDM (11). In molte di queste espressioni *rosso* assume il significato traslato di 'comunista, marxista, di sinistra' (11b). Alcune delle polirematiche verbali (11c) coinvolgono espressioni nominali (es. *tappeto rosso*) o preposizionali (es. *in rosso*) che sono a loro volta polirematiche a sé stanti.

- (11) a. semaforo rosso, capelli rossi, tappeto rosso, pianeta rosso, gigante rossa¹⁰, telefono rosso, conto in rosso, alga rossa
- b. brigatista rosso, cooperativa rossa, coop rossa, toghe rosse, terrorismo rosso, pericolo rosso, votare rosso, ecc.
- c. scattare l'allarme rosso, diventare rosso, avere l'occhio / gli occhi rossi, stendere il tappeto rosso, andare in rosso, passare col semaforo rosso
- d. da cartellino rosso.

Infine, per il termine *rosso* EXTra estrae un numero cospicuo di composti del tipo Colore+AGG e Colore+N:

⁹ Nome comune della specie *Delphinapterus leucas*, ma anche soprannome della Democrazia Cristiana.

¹⁰ Stella gigante di massa piccola o intermedia.

- (12) rosso fuoco, rosso sangue, rosso mattone, rosso fiammante, rosso fiamma, rosso pompeiano, rosso carminio, rosso rubino, rosso scarlatto, rosso vermiglio, rosso cupo, rosso bandiera, rosso acceso, rosso vino, rosso vivo.

Solo i primi 4 si trovano nel nostro campione, tutti gli altri sono collocati più in basso nella lista.

3.3.4 *Azzurro*

Delle 6 polirematiche registrate nel NDM e presenti nelle liste EXTra, 3 compaiono nel campione in esame (*maglia azzurra*, *principe azzurro*, *pesce azzurro*), mentre le altre 3 (*arma azzurra*, *camicia azzurra*, *telefono azzurro*) si trovano molto più in basso. Nel nostro sottoinsieme di dati troviamo anche polirematiche assenti nel NDM:

- (13) angelo azzurro¹¹, telefono azzurro, fiocco azzurro, cavaliere azzurro, balena azzurra¹², basco azzurro.

Molte delle stringhe che occupano le posizioni più alte nelle liste EXTra contengono *azzurro* come sinonimo di ‘italiano, nazionale, relativo alla nazionale italiana’, es.: *ct azzurro*, *squadra azzurra*, *nazionale azzurra*, *tecnico azzurro*, *panchina azzurra*, *ritiro azzurro*, *atleta azzurro*, ecc. Dalle liste possiamo recuperare 3 sole combinazioni del tipo Colore+N (solo la prima è presente nel campione):

- (14) azzurro cielo, azzurro mare, azzurro acqua.

3.3.5 *Verde*

Nel campione di dati qui esaminato troviamo le seguenti polirematiche già registrate nel NDM:

- (15) benzina verde, tavolo verde, raggio verde, tappeto verde, berretto verde, pollice verde, carta verde, tè verde, polmone verde, pepe verde.

Espressioni piuttosto comuni come *verde bottiglia*, *zona verde*, *numero verde* o *al verde* si trovano molto più in basso. Grazie all'estrazione con EXTra recuperiamo anche le seguenti espressioni non presenti nel NDM:

- (16) a. semaforo verde, luce verde, lista verde, libro verde, telefono verde, bandiera verde, panno verde, rivoluzione verde, verde età
b. deputato verde, senatore verde, capogruppo verde, eurodeputato verde, parlamentare verde, votare verde
c. dare il semaforo verde, dare la luce verde.

Come accade anche per altri lemmi, tra le stringhe verbali troviamo diversi sintagmi aventi come oggetto una polirematica nominale già individuata separatamente, ad esempio *numero verde* (*chiamare*, *istituire*, *attivare*).

Le liste EXTra contengono anche un discreto numero di composti con *verde* come testa (solo i primi 3 si collocano all'interno del nostro campione):

- (17) verde oliva, verde smeraldo, verde prato, verde acqua, verde bandiera, verde militare.

¹¹ Nome di un cocktail alcolico.

¹² Nome comune della specie *Balaenoptera musculus*, detta anche balenottera azzurra.

3.3.6 *Blu*

Nel campione di dati preso in esame sono presenti le seguenti espressioni già registrate nel NDM:

- (18) basco blu, zona blu, blu elettrico, casco blu, auto blu.

Tuttavia, notiamo che *casco blu* e *auto blu* non compaiono da soli, ma in unione con altri elementi (es. *casco blu francese*, *casco blu olandese*, *auto blu con autista*). Le prime occorrenze di *casco blu* da solo e *auto blu* da solo sono successive (posizioni 151 e 158, rispettivamente).

Nel campione sono presenti anche le espressioni in (19):

- (19) bandiera blu, fascia blu¹³, fifa blu, amianto blu¹⁴, balena blu¹⁵, periodo blu, pillola blu.

Per quanto riguarda i termini di colore composti, oltre al già citato *blu elettrico*, abbiamo altri 3 risultati (di cui solo il primo figura nel campione):

- (20) blu scuro, blu metallizzato, blu notte.

3.3.7 *Giallo*

Anche per *giallo* riportiamo le polirematiche presenti nel campione EXTra che sono già registrate nel NDM (21), e quelle invece non contenute nella risorsa lessicografica (22).

- (21) maglia gialla, cartellino giallo, fiamme gialle, bandiera gialla, stella gialla, pagine gialle, pericolo giallo, risotto giallo, metallo giallo, oro giallo.

- (22) romanzo giallo, febbre gialla, libro giallo, muso giallo, racconto giallo, zucca gialla.

Anche in questo caso i composti con *giallo* come testa sono in numero ridotto, ma tutti presenti nel campione:

- (23) giallo canarino, giallo limone, giallo oro, giallo paglierino.

3.3.8 *Grigio*

Sono solo 4 le polirematiche con *grigio* che EXTra riesce a estrarre, e si trovano tutte nel campione in esame:

- (24) lupo grigio, pantera grigia, materia grigia, eminenza grigia.

Abbiamo poi alcune polirematiche, tra cui figurano anche un paio di espressioni verbali, inedite rispetto al NDM:

- (25) a. zona grigia, fumata grigia, mercato grigio¹⁶, balena grigia¹⁷
b. avere i capelli grigi, vedere grigio.

¹³ Fascia oraria per il consumo dell'energia elettrica.

¹⁴ Tipo di amianto noto anche come crocidolite.

¹⁵ Altro nome per balena azzurra.

¹⁶ Compravendita di beni, a prezzi vantaggiosi, al di fuori dei normali canali di distribuzione (spesso tra paesi diversi).

¹⁷ Nome comune della specie *Eschrichtius robustus*.

Espressioni come *cielo grigio* e *giornata grigia*, che sono riportate molto in alto nelle liste EXTra, ci sembrano caratterizzabili più come combinazioni preferenziali che non come polirematiche.

All'interno del nostro campione troviamo anche i seguenti composti con *grigio* come testa:

- (26) grigio metallizzato, grigio fumo, grigio perla, grigio ferro, grigio topo, grigio piombo, grigio cielo.

3.3.9 Viola

Il NDM non registra nessuna polirematica con il lessema *viola*. EXTra estrae un numero non trascurabile di candidati (94), tuttavia nella maggior parte di queste combinazioni la forma *viola* si riferisce non al colore viola ma alla squadra di calcio *Fiorentina* (cfr. es. *tifoso viola*, *società viola*, *squadra viola*, *ultrà viola*, *tecnico viola*, *giocatore viola*, *in maglia viola*, *restare in viola*, ecc.).

3.3.10 Rosa

Per *rosa* EXTra estrae solo 83 stringhe, quindi l'intero set di dati è stato esaminato (come del resto per *viola*). Troviamo dunque tutte le polirematiche registrate nel NDM (27), ma anche qualche polirematica in più (28).

- (27) maglia rosa, cronaca rosa, foglio rosa, fenicottero rosa, vedere rosa.

- (28) a. romanzo rosa, pantera rosa, fiocco rosa, neorealismo rosa, apostrofo rosa, periodo rosa, spiaggia rosa
b. vedere rosa.

Accanto al più convenzionalizzato *romanzo rosa* troviamo anche *romanzetto rosa*, *rotocalco rosa*, *stampa rosa*, *letteratura rosa*, *favola rosa*, *giornale rosa*, *scrittore rosa*, *storia rosa*. Il fatto che *rosa* abbia assunto un significato traslato indipendente ('romantico, d'amore') aumenta dunque la combinabilità di questo termine (cfr. anche *scandalo rosa*, *triangolo rosa*).

I composti con *rosa* come testa ammontano invece a 4:

- (29) rosa confetto, rosa shocking, rosa pallido, rosa salmone.

3.3.11 Arancione

Le 13 stringhe estratte da EXTra per *arancione* contengono una sola combinazione di parole rilevante ai nostri fini, ovvero (*in*) *tuta arancione*, che può riferirsi sia alla tipica tuta da lavoro usata dagli operai (specialmente su strada o in cantiere), sia alla divisa da carcerato statunitense.

3.3.12 Marrone

EXTra non ha estratto alcuna stringa con il lessema *marrone*. Il dato quindi è pienamente in linea con il NDM, che non registra alcuna polirematica per questo termine.

4. Conclusioni

Lo studio quantitativo qui condotto, sebbene preliminare, ha mostrato il ruolo attivo dei termini di colore nelle polirematiche dell'italiano. L'ipotesi iniziale, che prevedeva un maggiore coinvolgimento dei termini di colore più frequenti – nonché di quelli più in alto nella scala implicazionale di Berlin & Kay (1969) – nella formazione di polirematiche, ha trovato conferma: stando ai dati fin qui analizzati, i termini *bianco*, *nero* e *rosso* sono largamente i più usati nella formazione di polirematiche, seguiti da *verde*.

Azzurro e *blu* compaiono in un numero minore di espressioni, che però si avvicina a quello di *verde* se considerati insieme. Anche *giallo* e *grigio* presentano una buona combinabilità, così come *rosa*.

Lo studio ha mostrato anche come l'estrazione semi-automatica di informazione combinatoria da corpora tramite metodi mirati possa contribuire notevolmente all'ampliamento delle nostre conoscenze nell'ambito della combinatoria lessicale: l'estrazione con EXTra ha infatti consentito di individuare un alto numero di polirematiche non registrate nel NDM, e questo nonostante l'analisi si sia concentrata solo su un campione di dati limitato rispetto a quello complessivo disponibile. Peraltro, il presente studio si è focalizzato sulle polirematiche, ma i dati a disposizione si prestano a un'analisi sulla combinatoria lessicale in genere (collocazioni e combinazioni preferenziali).

Infine, se da un lato corpora e tecniche computazionali possono aiutarci ad affinare le nostre analisi, dall'altro anche tali tecniche possono essere affinate grazie all'intuizione e alle osservazioni del linguista che ne fa uso. Dai dati EXTra, ad esempio, non emergono alcune combinazioni contenenti termini di colore che sono invece rilevanti per il nostro studio. Si tratta di costruzioni di 'somialianza' come quelle in (30), in cui il termine di colore è seguito da *come* e da un sintagma nominale:

(30) nero come la pece, bianco come la neve, rosso come il sangue.

Tali costruzioni sono state analizzate in Masini (2016, 2017), in relazione anche ai composti del tipo Colore+N di cui abbiamo parlato più sopra (es. *rosso sangue* vs. *rosso come il sangue*): entrambi infatti si basano su un meccanismo di comparazione con il colore tipicamente associato al referente del nome coinvolto. L'assenza di queste espressioni nei dati EXTra è dovuta al fatto che il pattern relativo a questa costruzione (che peraltro non si limita ai soli termini di colore, es. *vecchio come il mondo*) non figura tra le sequenze POS utilizzate in *CombiNet*, che potrebbero quindi essere arricchite, consentendo così l'estrazione di ulteriori combinazioni di parole significative.

Riferimenti bibliografici

- Attardi, Giuseppe & Felice Dell'Orletta. 2009. Reverse revision and linear tree combination for dependency parsing. *Proceedings of NAACL 2009*, 261-264.
- Baroni, Marco, Silvia Bernardini, Federica Comastri, Lorenzo Piccioni, Alessandra Volpi, Guy Aston & Marco Mazzoleni. 2004. Introducing the *la Repubblica* corpus: A large, annotated, TEI(XML)-compliant corpus of newspaper Italian. *Proceedings of LREC 2004*, 1771-1774.
- Berlin, Brent & Paul Kay. 1969. *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.
- Castagnoli, Sara, Gianluca Lebani, Alessandro Lenci, Francesca Masini, Malvina Nissim & Lucia Passaro. 2016. POS-patterns or syntax? Comparing methods for extracting Word Combinations. In Gloria Corpas Pastor (ed.), *Computerised and corpus-based approaches to phraseology: Monolingual and multilingual perspectives*, 101-114. Geneva: Tradulex.
- De Mauro, Tullio. 1999. Introduzione. In Tullio De Mauro (ed.), *Il grande dizionario italiano dell'uso*, vol. 1, VII-XLII. Torino: UTET.
- De Mauro, Tullio. 2005. *La fabbrica delle parole. Il lessico e problemi di lessicologia*. Torino: UTET.
- D'Achille, Paolo & Maria Grossmann. 2010. I composti aggettivo + aggettivo in italiano. In Maria Iliescu, Heidi M. Siller-Runggaldier & Paul Danler (eds.), *Actes du XXV^e Congrès International de Linguistique et de Philologie Romanes (Innsbruck, 3-8 septembre 2007)*, Tome VII, 405-414. Berlino: De Gruyter.
- D'Achille, Paolo & Maria Grossmann. 2013. I composti <colorati> in italiano tra passato e presente. In Emili Casanova Herrero & Cesáreo Calvo Rigual (eds.), *Actas del XXVI Congreso Internacional de Lingüística y de Filología Románicas (Valencia, 6-11 septiembre 2010)*, 523-537. Berlino: De Gruyter.
- Dell'Orletta, Felice. 2009. Ensemble system for Part-of-Speech tagging. *Proceedings of EVALITA 2009*.
- Grossmann, Maria. 1988. *Colori e lessico. Studi sulla struttura semantica degli aggettivi di colore in catalano, castigliano, italiano, romeno, latino ed ungherese*. Tübingen: Narr.
- Grossmann, Maria & Paolo D'Achille. 2016. *Italian colour terms in the BLUE area: Synchrony and diachrony*. In João Paulo Silvestre, Esperança Cardeira & Alina Villalva (eds.), *Colour and colour naming: Crosslinguistic approaches*, 21-50. Lisbona: Centro de Linguística da Universidade de Lisboa / Universidade de Aveiro.

- Lenci, Alessandro, Francesca Masini, Malvina Nissim, Sara Castagnoli, Gianluca Leboni, Lucia Passaro & Marco Senaldi (2017). *How to harvest word combinations from corpora: Methods, evaluation and perspectives*. Articolo proposto per la pubblicazione.
- Masini, Francesca. 2011. Polirematiche. In Raffaele Simone (ed.), *Enciclopedia dell'Italiano*, 1109-1112. Roma: Istituto dell'Enciclopedia Italiana G. Treccani.
- Masini, Francesca. 2016. *Morphological words and multiword expressions: Competition or cooperation?* Relazione presentata al 17° International Morphology Meeting (IMM17), Vienna, 18-21 febbraio 2016.
- Masini, Francesca. 2017. *Competition between morphological words and multiword expressions*. Articolo proposto per la pubblicazione.
- Passaro, Lucia & Alessandro Lenci. 2016. *Extracting terms with EXTra*. In Gloria Corpas Pastor (ed.), *Computerised and corpus-based approaches to phraseology: Monolingual and multilingual perspectives*, 188-196. Geneva: Tradulex.
- Voghera, Miriam. 2004. Polirematiche. In Maria Grossmann & Franz Rainer (eds.), *La formazione delle parole in italiano*, 56-69. Tübingen: Niemeyer.

Appendice: elenco delle polirematiche 'di colore' nel Nuovo De Mauro.

bianco	131	a pane bianco, abete bianco, acero bianco, acque bianche, agarico bianco, airone bianco maggiore, all'arma bianca, alzare bandiera bianca, andare in bianco, arma bianca, assegno in bianco, bandiera bianca, bianco d'uovo, bianco di zinco, bianco e nero, bianco e rosso, bianco mangiare, caffè bianco, caglio bianco, calore bianco, cambiale in bianco, camice bianco, carlina bianca, carne bianca, carta bianca, Casa Bianca, cedro bianco, cera bianca, cioccolato bianco, clown bianco, colletto bianco, cravatta bianca, cronaca bianca, cucitrice di, in bianco, da segnare col carbone bianco, delfinattero bianco, di punto in bianco, effluente bianco, elettrodomestici bianchi, far venire i capelli bianchi, farina bianca, fieno bianco, firma in bianco, firmare in bianco, firmare un assegno in bianco, folle bianca, frittura bianca, fumata bianca, gelso bianco, giglio bianco, girata in bianco, globulo bianco, golpe bianco, gorgonzola bianca, gramigna bianca, guelfo bianco, in bianco, in bianco e nero, lavagna bianca, lepre bianca, libro bianco, loto bianco d'Egitto, luce bianca, lupara bianca, lupino bianco, magia bianca, mal bianco, matrimonio bianco, matrimonio in bianco, mettere nero su bianco, monte bianco, morte bianca, mosca bianca, nana bianca, nebbiolo bianco, ninfea bianca, notte bianca, olio bianco, olmo bianco, omicidio bianco, ontano bianco, oro bianco, orso bianco, pane bianco, papavero bianco, partita in bianco, pastenula bianca, pepe bianco, perdite bianche, pernice bianca, pinot bianco, pioppo bianco, pizza bianca, poiiana dalla coda bianca, polvere bianca, rinoceronte bianco, risultato bianco, ruggine bianca, russo bianco, salice bianco, salsa bianca, sandalo bianco, scheda bianca, sciopero bianco, semestre bianco, semestre in bianco, semolino bianco, senape bianca, settimana bianca, smergo bianco, squalo bianco, stella nana bianca, strada bianca, tartufo bianco, telefoni bianchi, terra bianca, terrore bianco, tonno bianco, topo bianco, tratta delle bianche, trattare con i guanti bianchi, treno bianco, trifoglio bianco, trombo bianco, vedova bianca, vetro bianco, vinificazione in bianco, vino bianco, vite bianca, voce bianca, volpe bianca
nero	91	acque nere, angelo nero, anima nera, aristocrazia nera, bandiera nera, basco nero, bestia nera, bianco e nero, borsa nera, borsaro nero, brigate nere, buco nero, caffè nero, camicia nera, carne nera, cavolo nero, continente nero, corallo nero, cornacchia nera, corpo nero, cravatta nera, cronaca nera, di umore nero, effluente nero, fagiolo nero, fare nero, fiamme nere, fondo nero, fumata nera, gelso nero, germano nero, giovedì nero, guelfo nero, in bianco e nero, in nero, lava nera, lavoro nero, libro nero, lista nera, magia nera, maglia nera, mano nera, marea nera, mercato nero, messa nera, mettere nero su bianco, morte nera, musulmano nero, nero di seppia, nero fumo, nobiltà nera, olio nero, onda nera, ontano nero, oro nero, orso nero, pane nero, pantera nera, papa nero, pece nera, pecora nera, penna nera, pepe nero, perla nera, peste nera, picchio nero, pinot nero, pioppo nero, piovanello pancia nera, polenta nera, porcino nero, pozzo nero, punto nero, ribes nero, rinoceronte nero, sambuco nero, scatola nera, senape nera, Settembre Nero, storno nero, tartufo nero, tè nero, umore nero, umorismo nero, uomo nero, vedere nero, vedere tutto nero, vedova nera, vernaccia nera, vino nero, vite nera

rosso	75	a luci rosse, abete rosso, acero rosso, airone rosso, armata rossa, bandiera rossa, barbabietola rossa, basco rosso, bianco e rosso, bietolone rosso, brigate rosse, calore rosso, camicia rossa, canguro rosso, carne rossa, cartellino rosso, cavolo rosso, corallo rosso, Croce Rossa, Croce Rossa Internazionale, Croce Rossa Italiana, disco rosso, fenicottero rosso, fiamme rosse, filo rosso, formica rossa, globulo rosso, guardia rossa, in rosso, khmer rosso, lattuga rossa, libretto rosso, linea rossa, lupino rosso, mezzaluna rossa, mirtillo rosso, nana rossa, ocras rossa, oro rosso, pagello rosso, panda rosso, papa rosso, partito rosso, passaporto rosso, pelle rossa, pepe rosso, perdilegno rosso, perdite rosse, pesce rosso, picchio rosso minore, pimento rosso, pizza rossa, porfido rosso antico, primula rossa, radica rossa, radicchio rosso, ragnetto rosso, rana rossa, ribes rosso, rosso d'uovo, salice rosso, sandalo rosso, soccorso rosso, spostamento gravitazionale verso il rosso, stella gigante rossa, stella rossa, telefono rosso, terra rossa, trifoglio rosso, trombo rosso, vedere rosso, vinificazione in rosso, vino rosso, volante rossa, volpe rossa
verde	44	al verde, balletto verde, basco verde, benzina verde, berretto verde, bollino verde, carta verde, cavalletta verde, colombina verde, Croce Verde, disco verde, fiamme verdi, frittata verde, in verde, insalata verde, lira verde, lui verde, moneta verde, numero verde, onda verde, ontano verde, pepe verde, picchio verde, pigna verde, pollice verde, polmone verde, porfido verde antico, raggio verde, rana verde, rana verde maggiore, rana verde minore, salsa verde, tappeto verde, tavolo verde, tè verde, valuta verde, vedere i sorci verdi, verde acido, verde antico, verde bottiglia, verde mare, verde penicillina, verde rame, zona verde
giallo	29	ambra gialla, bandiera gialla, cartellino giallo, ditola gialla, farina gialla, fiamme gialle, genziana gialla, giallo di stagno, giallo psicologico, in guanti gialli, ladro in guanti gialli, lupino giallo, maglia gialla, margherita gialla, metallo giallo, ocras gialla, oro giallo, pagine gialle, pan giallo, pane giallo, pericolo giallo, pioggia gialla, radica gialla, risotto giallo, semolino giallo, sindacato giallo, stampa gialla, stella gialla, trifoglio giallo delle sabbie
azzurro	19	arma azzurra, balenottera azzurra, camicia azzurra, corallo azzurro, croce azzurra, fiamme azzurre, ghiandaia azzurra, lupino azzurro, maglia azzurra, morbo azzurro, pagro azzurro, palma azzurra, parco azzurro, partito azzurro, passiflora azzurra, pesce azzurro, principe azzurro, sangue azzurro, telefono azzurro
grigio	18	ambra grigia, canguro gigante grigio, cornacchia grigia, eminenza grigia, foca grigia, grigio isabella, lega grigia, letteratura grigia, lupo grigio, materia grigia, materiale grigio, moretta grigia, notidano grigio, orso grigio, pantera grigia, pinot grigio, squalo grigio, tignola grigia della farina
blu	14	auto blu, bambino blu, basco blu, blu di metilene, blu elettrico, casco blu, colletto blu, mirtillo blu, morbo blu, parco blu, sangue blu, tuta blu, uomo blu, zona blu
rosa	11	balletto rosa, cartolina rosa, colletto rosa, cronaca rosa, fenicottero rosa, foglio rosa, maglia rosa, rosa antico, salsa rosa, telefono rosa, vedere rosa