

Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN
TRADUZIONE, INTERPRETAZIONE E
INTERCULTURALITÀ

Ciclo XXVIII

Settore Concorsuale di afferenza: 10/H1 - LINGUA, LETTERATURA E CULTURA
FRANCESE

Settore Scientifico disciplinare: L-LIN/04 – LINGUA E TRADUZIONE – LINGUA
FRANCESE

**APPROCCIO LINGUISTICO E *CORPUS-DRIVEN* AL PROVERBIO
ITALIANO E FRANCESE: ALLA RICERCA DELLA FORMA PERDUTA**

Presentata da: Vincenzo Lambertini

Coordinatore Dottorato

Prof.ssa Raffaella Baccolini

Relatore

Prof.ssa Licia Reggiani

Esame finale anno 2016

A Nadia

Indice

Introduzione	4
1. La linguistica del proverbio	7
1.1. Definizione di proverbio	9
1.1.1. L'etichetta "proverbio": una costante non universale.....	9
1.1.2. Condizioni necessarie del proverbio	13
1.1.3. Espressioni idiomatiche	16
1.2. Proverbi ed espressioni idiomatiche: un confronto sintattico	19
1.3. Proverbi ed espressioni idiomatiche: un confronto semantico	26
1.3.1. Una risposta: la vera differenza tra proverbi ed espressioni idiomatiche ...	27
1.4. Verso un nuovo paradigma definitorio di proverbio	30
1.4.1. Frase <i>ON-sentencieuse</i>	30
1.4.2. Frasi generiche	33
1.4.3. Proverbi e denominazione	34
1.4.4. La parabola del proverbio	40
1.5. Proverbi, <i>déproverbialisation</i> e <i>détournement</i>.....	42
1.6. Funzionamento dei proverbi nel discorso.....	46
2. Linguistica dei corpora e proverbi	50
2.1. Una scelta metodologica, ma non solo	50
2.1.1. Nascita e sviluppo della linguistica dei corpora.....	51
2.1.2. Corpora linguistici.....	53
2.1.3. Tipi di corpora	56
2.2. L'analisi dei proverbi in corpora linguistici e altri approcci.....	57
2.3. Caratteristiche dei corpora WaCky. Focus sul corpus itWaC.....	60
2.4. Perché itWaC e frWaC	63
2.5. Strumenti di consultazione dei corpora WaCky	67
2.6. Linguistica dei corpora e paremiologia	67
2.7. Verso un approccio <i>corpus-driven</i> all'analisi paremiologica	69
2.8. Esperienze di applicazione della linguistica dei corpora alla fraseologia.....	72
2.9. Come individuare automaticamente proverbi in corpora?	75
2.9.1. Tre dizionari di proverbi italiani a confronto	79
2.9.2. Primo tentativo fallito	81

2.9.3. Marcatori di proverbio e secondo tentativo.....	83
3. Estrazione automatica di proverbi italiani e francesi.....	89
3.1. Percorso seguito nell'individuazione dei proverbi da analizzare.....	89
3.2. Vaglio dei risultati.....	92
3.2.1. "Proverbi non italiani", "proverbi non francesi".....	92
3.2.2. Considerazioni sui "proverbi non italiani / non francesi".....	93
3.2.3. "Interesse storico-folkloristico".....	104
3.2.4. Proverbi appartenenti a opere letterarie anteriori al XX secolo.....	106
3.2.5. "No proverbio".....	107
3.2.6. "Ripetizioni".....	109
3.3. Analisi quantitativa delle concordanze validate e di quelle eliminate....	110
3.4. Natura dei proverbi individuati.....	113
3.5. Scelta delle informazioni da mantenere per ogni proverbio reperito....	114
3.6. Importazione dati e costruzione della tabella di lavoro in Excel.....	119
3.7. Occorrenze dei proverbi introdotti dalla parola <i>proverbio</i>.....	124
3.8. Liste di frequenza di proverbi: i frutti del nostro lavoro.....	137
4. Analisi di proverbi italiani e francesi.....	139
4.1. Scelta dei proverbi da cercare e analizzare in itWaC e frWaC.....	139
4.2. Come cercare i proverbi scelti nei due corpora.....	145
4.3. Fase di ricerca di proverbi senza l'uso dei marcatori.....	155
4.4. Altri marcatori di proverbio.....	171
4.5. Analisi del nostro marcatore di proverbio e riflessioni sulle risorse dedicate ai proverbi.....	173
4.6. Proverbi e variabilità.....	175
4.7. Ulteriori piste di ricerca.....	185
4.7.1. Proverbi e dizionari: come migliorare questo binomio.....	188
4.7.2. Variabilità e semantica dei proverbi.....	189
Conclusioni.....	191
Bibliografia.....	195
Corpora utilizzati: itWaC e frWaC.....	201
APPENDICE.....	202
ALLEGATO 1: Lista di frequenza dei proverbi reperiti con approccio <i>corpus-</i> <i>driven</i> in itWaC.....	202

ALLEGATO 2: Lista di frequenza dei proverbi reperiti con approccio <i>corpus-driven</i> in frWaC	228
ALLEGATO 3: <i>Database</i> di tutti i “proverbi adeguati all’analisi” reperiti in itWaC	
ALLEGATO 4: <i>Database</i> di tutti i “proverbi adeguati all’analisi” reperiti in frWaC	

Introduzione

La presente Tesi di Dottorato si sviluppa attorno a un tema spesso relegato a una posizione di secondo piano, ma che riveste, in realtà, un grande interesse linguistico e culturale: il *proverbio*. Si vedrà, nel corso della Tesi, che il proverbio, essendo spesso considerato nella sua accezione più folklorica e tradizionale, appare come un oggetto ormai in via di estinzione, destinato a tramontare come le società del passato a cui sembra riferirsi o in cui si è sviluppato.

Alla base di questo lavoro c'è una consapevolezza diametralmente opposta: il proverbio viene ancora oggi usato e, sebbene abbia una frequenza d'uso poco elevata, la sua enunciazione non passa inosservata. Quando in una conversazione si cita un proverbio, i partecipanti riconoscono che si tratta di un enunciato che condensa un significato molto più profondo. Allo stesso tempo, il proverbio stesso rinsalda i legami esistenti tra i partecipanti e la propria comunità linguistico-culturale, il che si può rivelare un'arma a doppio taglio: infatti, se, da un lato, consolida il senso di appartenenza dei partecipanti alla stessa comunità, dall'altro rischia di emarginare chi non fa parte della comunità.

Questo è il caso tipico di chiunque si relazioni con parlanti di lingue e culture diverse. In particolare, coloro che si trovano a mediare tra lingue e culture, come gli interpreti o i traduttori, vivono sulla propria pelle la difficoltà di gestire situazioni problematiche o perfino conflittuali, causate dalla presenza mera di un proverbio, che può risultare di difficile comprensione, interpretazione o traduzione.

In un certo senso, questa Tesi parte proprio dall'impotenza e dal disagio che possono provare i professionisti delle lingue e, in particolare, gli interpreti, nel caso in cui venga citato un proverbio a loro ignoto oppure non sufficientemente conosciuto nelle due lingue. Certo, la prima arma di difesa consiste nell'aggirare l'ostacolo, ma non sempre tale soluzione è in grado di contenere i danni.

Come si vedrà, nel corso della trattazione in oggetto i riferimenti all'ambito dell'interpretazione e della traduzione saranno alquanto scarsi. Non si analizzeranno dati relativi a eventi mediati da interpreti, né tantomeno traduzioni di testi in cui compaiono proverbi.

Al contrario, si cercherà di impostare correttamente l'analisi attorno ai proverbi, non dimenticando mai il nostro punto di partenza, rappresentato da un problema che si riscontra nella pratica della traduzione e dell'interpretazione.

Per questa ragione, ci si concentrerà dapprima sulla linguistica dei proverbi, al fine di individuare gli strumenti adeguati per analizzare correttamente il proverbio, per poi passare a una parte di analisi di dati, reali e autentici, estratti da due corpora di grandissime dimensioni: il corpus itWaC e il corpus frWaC. Le basi teoriche, pertanto, saranno necessarie per sviluppare l'intera ricerca condotta su dati reali.

Non mancheranno, in questo lavoro, frequenti riflessioni su quanto ancora ci sia da compiere in questo ambito di ricerca e su quanto lo studio del proverbio si sia basato per troppo tempo più su preconcetti o teorie create *ad hoc* che sull'osservazione dei dati reali.

Tale ricerca vuole essere un primo passo verso un approccio allo studio del proverbio più realistico e maggiormente legato alle esigenze di coloro che si dovranno misurare, prima o poi, con esso. Non a caso, si cercherà di mostrare come funzionano i proverbi in contesti autentici, senza escludere a priori i comportamenti potenzialmente problematici come quelli della presenza *proverbi modificati*, ma tentando sempre di trovare una risposta ai nostri interrogativi sulla base dell'osservazione dei dati reali e dei *pattern* evidenziati. Tenteremo altresì di mostrare quanto la percezione del proverbio (per la quale abbiamo coniato il sintagma *competenza paremiologica*) da parte del parlante comune possa divergere rispetto a quanto comunemente si asserisce in paremiologia. Ciò non significa in alcuna maniera che la competenza paremiologica comune debba sostituirsi ai principi della paremiologia, semmai, al contrario, impone una riflessione sul perché si possa verificare un divario tra quanto si consideri comunemente proverbio e quanto sia effettivamente proverbio. Tuttavia, non possiamo nemmeno censurare quanto succede nella comunicazione solo perché questo si pone in contrasto con quanto accade in paremiologia. A titolo di esempio, se, come si vedrà in § 4.5, un dizionario non attesta un proverbio che, invece, ha una frequenza d'uso molto elevata e soddisfa i criteri necessari per essere considerato come tale, in questo caso ci si deve interrogare sul perché una risorsa che dovrebbe fornire un aiuto si riveli inutile.

Prima di passare alla Tesi vera e propria, è opportuno sottolineare che questo lavoro, data la complessità, è stato possibile grazie al sostegno di esperti che, a diverso titolo, hanno accettato di offrire il proprio contributo.

Si ringrazia, in maniera particolare, il Professor Michele Prandi per i consigli forniti sulla linguistica e sulla linguistica del proverbio e delle espressioni idiomatiche. Un sincero ringraziamento va anche alla Professoressa Silvia Bernardini, per i suggerimenti dispensati su come sviluppare la parte relativa allo studio dei proverbi attraverso la *corpus linguistics*.

Molte delle idee emerse su come procedere nello studio sono scaturite in seguito a diversi colloqui precedenti all'inizio del presente progetto di ricerca, intrattenuti con i Professori Paolo Scampa e Franco Bertaccini, che non si sono mai rifiutati di elargire critiche e consigli.

Come non ringraziare, poi, il Professor Rui J. B. Soares, presidente dell'AIP-IPA (*Associação Internacional de Paremiologia / International Association of Paremiology*) per l'appoggio sempre mostrato e per il desiderio di far conoscere ai membri dell'associazione e ai partecipanti dei convegni di paremiologia che si tengono annualmente a Tavira, in Portogallo, la ricerca in oggetto. Il suo appoggio è andato oltre ogni ragionevole limite, mostrando costantemente un entusiasmo unico nei confronti di questa ricerca. Il ringraziamento si estende a tutti i membri dell'Associazione, che hanno sempre mostrato un forte interesse nei confronti di questa ricerca, nonostante le puntuali divergenze di opinioni.

Un doveroso ringraziamento va alla Professoressa Licia Reggiani per aver creduto fin dall'inizio in questa ricerca. A lei va, inoltre, il mio ringraziamento personale per la piena e incondizionata fiducia accordatami in ogni momento.

Infine, come ringraziare tutti coloro che mi hanno aiutato e sostenuto nei momenti più difficili, senza i quali questa Tesi non sarebbe stata possibile.

1. La linguistica del proverbio

L'approccio alla ricerca dei proverbi è molto spesso di tipo *top-down*, ovvero si parte da una serie di teorie sostenute da un'impalcatura di argomentazioni più o meno valide e condivise, per poi cercare un loro riscontro in esempi reperiti di volta in volta in opere letterarie, nella stampa o nella pubblicità. Si ha, dunque, l'impressione che si conosca già molto, se non tutto, sui proverbi. Probabilmente, questo dipende altresì dalla consapevolezza che i proverbi facciano parte della storia dell'umanità e che da sempre abbiano destato un certo interesse, da un lato per i loro contenuti di validità generale e, dall'altro, per la loro valenza culturale e folkloristica. Ciononostante, a un'analisi attenta della letteratura paremiologica, ci si renderà conto di quanto le teorie oggi esistenti siano, in realtà, contraddittorie e lacunose, segnatamente per quanto attiene al funzionamento linguistico e semantico del proverbio.

In un articolo pubblicato nel 2005 sul cinquantatreesimo numero della rivista *Linx*, Jean-Claude Anscombre (2005: 17) ritiene che l'influenza delle antiche grammatiche, che avevano spesso una visione molto normativa della lingua, anzi del "*beau parler*", abbia fatto sì che, fino a tempi recenti, i fenomeni sentenziosi venissero relegati a una categoria "*marginale et marginalisée*", marginale ed emarginata, appunto. Tali fenomeni, assieme ai proverbi, che ricoprono un ruolo centrale nell'ambito della presente ricerca, sono stati considerati in passato devianze rispetto alla normalità (o normatività?) della lingua, alla stregua delle interiezioni e delle onomatopee, e per questo non degne di essere oggetto di studi linguistici, ma semmai soltanto di ricerche in campo antropologico e folkloristico. "*La connotation négative attachée à folklore – en tant que mot de la langue et non en tant que notion – montre bien le peu d'estime (linguistique) attachée à cette sorte de manifestation*" (Anscombre, 2005: 17).

Si può pertanto affermare che vi sia una certa scarsità di studi linguistici in ambito paremiologico, benché sia errato ritenere che tale campo non desti l'interesse dei ricercatori. Al contrario, se consideriamo filoni di studi maggiormente legati al folklore, alla letteratura e all'antropologia, notiamo che le ricerche in ambito paremiologico abbondano. È, dunque, indubbio che tutt'ora si producano numerosi studi, più o meno legati alla linguistica, che riguardano i proverbi. Basti pensare ai molti lavori che sono stati e vengono tutt'ora prodotti in ambito paremiologico anche a livello internazionale. In Spagna, ricordiamo la rivista

*Paremia*¹, istituita dalla Professoressa Julia Sevilla Muñoz dell'Università Complutense di Madrid (Dipartimento di Filologia Francese) nel 1993. Per il Portogallo, citiamo l'intensa attività dell'Associazione Internazionale di Paremiologia (AIP-IAP)², con sede a Tavira. Tale associazione mira a stimolare la cooperazione internazionale nell'ambito della paremiologia e delle discipline ad essa legate, ad aiutare i giovani ricercatori attivi nell'ambito della paremiologia e della conservazione del patrimonio culturale intangibile, nonché a organizzare convegni internazionali e interdisciplinari di paremiologia. In area francofona, degni di nota sono i contributi di studiosi di proverbi come Jean-Claude Anscombe, Georges Kleiber, Silvia Palma, Laurent Perrin, Irène Tamba, Mirella Conenna, tutti autori i cui lavori sono raccolti nel centotrentanovesimo numero della rivista *Langages* (2000), un fascicolo di grande importanza per le basi metodologiche degli studi paremiologici. Oltre a questi autori, sono da ricordare altresì Yves-Marie Visetti e Pierre Cadiot che nel 2006 hanno pubblicato la monografia *Motifs et proverbes* (2006).

Per quanto attiene al nostro Paese, si nota che gli studi sui proverbi sono numerosi ma non molto recenti. Il contributo di Luciano Agostiniani (1978), "Semantica e referenza del proverbio", appartenente al Volume LXIII, fascicolo 1-2, dell'*Archivio Glottologico Italiano*, è sicuramente una pietra miliare in questo campo, senza dimenticare l'opera a cura di Cristina Vallini, *La pratica e la grammatica. Viaggio nella linguistica del proverbio* che raccoglie una serie di contributi che spaziano dalla storia del proverbio ad analisi più propriamente linguistiche relative ai proverbi, passando per la traduzione del proverbio. Il fatto che i principali studi paremiologici, o almeno quelli più legati alla linguistica del proverbio, siano abbastanza datati apre la questione dell'attualità dello studio dei proverbi. Si ha quasi l'impressione che essi costituiscano una materia di scarso interesse, il che dovrebbe interrogare la comunità scientifica che opera in ambito paremiologico circa le ragioni di ciò. Ad esempio, sarebbe tutt'altro che superfluo comprendere appieno l'utilizzo del proverbio e individuarne la frequenza d'uso, producendo anche considerazioni di tipo diacronico. Purtroppo, si ha la sensazione che il proverbio non rivesta più, per la linguistica italiana, quell'interesse che aveva in passato o che ha tutt'ora in altri paesi.

Come si vedrà anche successivamente, la penuria di studi linguistici sui proverbi costituisce un notevole ostacolo, specialmente in ragione dell'incertezza nel definire che cosa

¹ Per maggiori informazioni, si rimanda al sito internet: www.paremia.org.

² Per maggiori informazioni, si rimanda al sito internet: www.aip-iap.org.

sia un proverbio e nel differenziarlo dalle altre forme sentenziose. È ragionevole credere che tale scarsità di studi sia dovuta ad alcune opinioni riguardanti il proverbio, diffuse anche in ambito accademico. La prima di queste è che l'uso dei proverbi vada via via scomparendo. La seconda, direttamente legata alla prima, è che non si creino più proverbi. La terza, è che il proverbio sia un oggetto polveroso che, oltre ad essere passato di moda, è stato oggetto di troppe ricerche da cui è emerso tutto (o quasi) quello che c'era da sapere. In realtà, si vedrà da subito che del proverbio si ignora ancora molto e si smentiranno, nel corso della presente Tesi di Dottorato, le tesi appena esposte circa la scarsa frequenza d'uso, la scarsa produttività e la mancanza di creazione del proverbio.

1.1. Definizione di proverbio

1.1.1. L'etichetta “proverbio”: una costante non universale

Cerchiamo di capire in che cosa consista l'etichetta “proverbio”. Diatopicamente e diacronicamente, il proverbio ha assunto numerosi significati, ma non sempre è stato necessario attribuirgli un'etichetta. Se si assume una prospettiva storica, ci si rende conto che, sebbene i proverbi siano sempre esistiti, non è stata sempre assegnata loro un'etichetta. Da un lato, i proverbi sono antichissimi: pensiamo alle raccolte di proverbi sumerici³, che ancora oggi sorprendono i lettori per la sconvolgente attualità delle tematiche trattate, oppure agli antichi egizi, per i quali i proverbi venivano designati con il termine *sebayt*, che significa “insegnamento”. Allo stesso modo, presso gli aramei e gli ebrei i proverbi erano detti “parola”, intendendo la parola di un saggio. Per i greci, i termini *gnômê* (pensiero) e *paroemia* (istruzione) ricoprono le nozioni di proverbio, sentenza, massima, adagio, precetto, ecc. (Maloux, 1990: VII). Mantenendo una prospettiva storica, si nota che presso le civiltà greca e latina i termini *παροιμία* e *proverbium*, che designavano appunto i proverbi, sono comparsi soltanto in epoca classica e non erano, dunque, presenti in epoca arcaica. Gli autori arcaici, infatti, per designare i proverbi, in mancanza di tecnicismi, usavano iperonimi come *parola*, *discorso*, *racconto*, *insegnamento*. Questo probabilmente era dovuto al fatto che i proverbi, appartenendo alla tradizione orale, e quindi essendo semplicemente pronunciati, non imponevano l'uso di un termine metalinguistico per essere identificati. I termini *παροιμία* e *proverbium* compaiono soltanto quando, figurando nei testi scritti, i proverbi diventano oggetto di riflessioni linguistiche e, soprattutto, oratorie. Il tecnicismo diventa, dunque, indice,

³ Per un approfondimento, si veda Silvestri (1989).

e al contempo conseguenza, della forza oratoria di autorevolezza documentaria acquisita dai proverbi (Pocetti, 1989).

Assumendo un'ottica diatopica, si nota che molte lingue non posseggono ancora oggi alcun termine esclusivamente destinato ad indicare i proverbi. Si pensi, a mero titolo di esempio, che nella lingua africana IoDagaa il termine *yelbie* (*fanciullo* o *seme di parola*) può designare, in base ai contesti, qualsiasi segmento di discorso, dal morfema, alla frase, al verso di un canto, fino al proverbio e a un tema generale (Visetti & Cadiot, 2006: 13, nota 1).

Non è, infatti, detto che il proverbio sia necessariamente una frase. Non solo diatopicamente, come nell'esempio appena citato della lingua africana IoDagaa, ma anche diacronicamente il proverbio ha superato frequentemente la mera unità frastica per farsi testo. Charlotte Schapira (2000: 85 e 1999: 68) evidenzia come la quartina riportata da Proust in *A la recherche du temps perdu* fosse in realtà un proverbio:

- (1) - Mangeons mon pain,
- Je le veux bien.
- Mangeons le tien.
- Je n'ai plus faim.

Si tratta a tutti gli effetti di un testo coerente e coeso: sebbene non siano presenti ponti anaforici che esplicitino le relazioni tra gli enunciati, la coesione testuale è comunque garantita dalla presenza del pronome “*le*” che riprende anaforicamente il sostantivo “*pain*”. Ad ogni modo, come si vedrà, è alquanto frequente ritrovare nei proverbi una sintassi “slegata per l'assenza di elementi connettivi” (Cristilli, 1989: 196).

Altri esempi di proverbi che superano l'unità frastica si possono ritrovare in Anscombe (2000: 13), che ritiene che sia più corretto parlare di *discorsi* a riguardo dei proverbi anziché di *frasi*. Egli considera che non vi sia nulla, in un'ottica sincronica, diacronica o culturale, se non una mera tendenza attuale, che ci spinga a considerare i proverbi come frasi, in virtù della frequente possibilità che essi vadano oltre la semplice unità frastica. Così, anche in Anscombe (2000: 13) ritroviamo esempi di proverbi più simili a testi che a frasi:

- (2) Blanc sur rouge
Tout bouge
Rouge sur blanc
Tout fout le camp

- (3) S'il pleut à la Saint-Médard
Il pleut quarante jours plus tard
Mais vient le bon Saint-Barnabé
Qui peut encore tout arranger

Esempi di questo tipo non mancano, specialmente se pensiamo alle raccolte di proverbi medievali in cui è difficile distinguere il proverbio dal testo in cui appare. Anscombe (2000: 13) cita la raccolta *Le proverbe au vilain* che annovera veri e propri testi al cui interno si possono individuare versi che potrebbero assomigliare ai nostri proverbi, ma senza avere la certezza di che cosa all'epoca potesse essere considerato proverbio. Inoltre, basti pensare a testi ancora oggi molto noti, come le *Favole* di De La Fontaine, che all'inizio erano *in toto* proverbi e poi hanno solo successivamente dato vita a proverbi condensati in frasi.

I proverbi che abbiamo ottenuto interrogando il corpus itWaC,⁴ nella fase della tesi che si presenterà successivamente, mostrano quanto permangano ancora oggi proverbi composti da due frasi. Pensiamo ai seguenti esempi:

- (4) a. Errare è umano, perseverare è diabolico
b. Natale con i tuoi, Pasqua con chi vuoi
c. Fidarsi è bene, non fidarsi è meglio

Se consideriamo i testi e i discorsi come “sequenze più o meno lunghe di enunciati” (Prandi, 2006: 171), possiamo affermare che i proverbi dell'esempio (4), una volta utilizzati nella comunicazione scritta od orale, siano dei testi o dei discorsi. Notiamo, infatti, che i proverbi in (4) sono formati da due frasi indipendenti, che non presentano tra loro rapporti grammaticali, ma che mettono semplicemente in relazione i propri contenuti. Esse, infatti, potrebbero essere separate graficamente con un punto oppure scritte in due righe distinte, come se fossero versi di una poesia. Una prova dell'autonomia delle frasi che compongono ognuno dei proverbi in (4) risiede nella possibilità di citare anche solamente una parte del proverbio, indistintamente la prima o la seconda⁵, come avviene negli esempi in (5), reperiti grazie alle ricerche sul corpus itWaC.

⁴ Per maggiori dettagli, si veda § 2.3.

⁵ Si tornerà successivamente a trattare (cfr. § 4) della frequente ed effettiva possibilità di citare solo in parte i proverbi o di modificarli. Si tratta di fenomeni molto frequenti che possono essere racchiusi sotto il grande cappello della *déproverbialisation* e del *détournement*.

- (5) a. “Come è noto perseverare è diabolico ma non evidentemente per il ministero. Ciò spiega perché ad oltre dieci anni dalla entrata in vigore della legge quadro e ad oltre 20 dalla legge sul mare si continui genericamente a parlare di aree marine protette senza alcuna distinzione tra parchi e riserve (...).” [da: <http://comunic.it/PARKS/parco.conero/notiziario/2004-luglio/3.1-coste.html>]
- b. “Pasqua con chi vuoi... o con chi rimane. Ciao amici belli, mi sento un po' come l'asinello triste ih-oh perchè causa febbre ostinata di Bianca (con tutte le lagne connesse) non possiamo andare al mare né fare nient'altro per Pasqua.” [da: <http://coffeebreak.ubiks.it/index.php/20050324>]
- c. “Ho acquistato il majesty nel aprile 2001 adesso nel mese di agosto 2003 a circa 23.500 km sentivo un rumore alquanto strano, proveniente dal motore, come di una mancata lubrificazione di organi meccanici ma non mi preoccupavo perchè avevo fatto un tagliando presso concessionario a 21.000 km, evidentemente non fidarsi è meglio e recandomi dal concessionario dopo aver verificato che il motore era praticamente senza olio, provvedeva ad aggiungerne altro.” [da: <http://www.cyberscooter.it/sx/posta03b.htm>].

Le frasi eliminate dai tre proverbi non impediscono la comprensione del testo, né tantomeno il corretto riconoscimento del proverbio stesso. Immaginando la lingua del futuro, potremmo dire che, teoricamente, nulla impedisce a questi tre proverbi di affermarsi unicamente in versione abbreviata. Come sempre, sarà l'uso e l'evoluzione naturale della lingua a decidere come e quali saranno i proverbi di domani.

Ricapitolando, è giusto ritenere che i proverbi siano frasi. Statisticamente, i dati in nostro possesso indicano che la maggioranza dei proverbi italiani e francesi oggi esistenti e utilizzati nella comunicazione sono composti da una sola frase. Seguono, poi, in percentuale nettamente inferiore, proverbi composti da almeno due frasi ma con relazioni anaforiche che le legano, quasi a formare un'unità. Infine, una percentuale molto bassa è composta da proverbi formati da più frasi che non presentano relazioni anaforiche, il che contribuisce alla loro scarsa coesione testuale, di cui si parlerà successivamente.

Tuttavia, gli esempi (1), (2), (3), e (4) impongono di tenere in considerazione la natura testuale dei proverbi, i quali possono essere veri e propri testi coerenti e, sebbene non nella totalità dei casi, coesi. La conclusione che è possibile trarre è che esiste una soglia minima al di sotto della quale i proverbi non sono accettabili: i proverbi non possono essere meno di una

frase (ossia, non possono essere costituenti di frasi come sono, ad esempio, le espressioni idiomatiche), ma possono essere un insieme di frasi, vale a dire un testo. Si approfondirà questo punto in § 1.1.2.

Questa breve nota si è resa necessaria per evidenziare l'approccio scelto nell'ambito della presente Tesi di Dottorato, la quale è volta a studiare i proverbi effettivamente utilizzati oggi in lingua italiana e francese. Per ragioni di praticità e di frequenza statistica si parlerà, nel corso della presente ricerca, di frasi ed enunciati in riferimento ai proverbi, sebbene sia evidente che questi possano essere, più frequentemente di quanto si possa credere, veri e propri testi.

1.1.2. Condizioni necessarie del proverbio

Non di raro accade che, in numerosi studi in ambito paremiologico, si insista su quanto sia arduo determinare le condizioni necessarie del proverbio. Prendiamo, come esempio, l'incipit del paragrafo dedicato ai proverbi e alle forme proverbiali dell'articolo di Anscombe (1994: 95), "*Proverbes et formes proverbiales : valeur évidentielle et argumentative*", il quale avverte la necessità di sottolineare da subito la mancanza di una definizione accettata di proverbio: "*D'entrée, on se heurte à l'obstacle (souvent signalé) de la définition du proverbe*". La difficoltà di avere una definizione condivisa ed esaustiva, tale da descrivere tutte le proprietà del proverbio, è stata messa in luce anche da Cardona (2010: 166), il quale ricorda come anche Archer Taylor negasse "l'utilità e perfino la possibilità di definizioni valide del proverbio, che possiederebbe una sua «qualità comunicabile»". Molto arguta ma scarsamente precisa risulta la modalità con cui Luciano Agostiniani individua i proverbi da prendere in considerazione nell'articolo "Semantica e referenza nel proverbio" (Agostiniani, 1978). Egli indica che saranno "considerati come proverbi tutti e solo quelli che come tali sono stati appresi da chi scrive e/o come tali sono stati forniti dagli informatori dell'API⁶". Un criterio evidentemente poco scientifico e troppo soggettivo, ma soprattutto che non fornisce strumenti adeguati per assolvere tale compito.

Uno dei primi requisiti ritenuti indispensabili da molti autori affinché una frase si faccia proverbio è la concisione. In altre parole, i proverbi sarebbero *in primis* frasi brevi e concise. Nemmeno questo, tuttavia, è un criterio chiaro e obiettivo. Come si può calcolare la lunghezza di una frase e decidere quando sia breve e concisa? Quale dovrebbe essere il metro

⁶ Atlante Paremiologico Italiano.

di paragone? Inoltre, si è dimostrato in § 1.1.1 che il proverbio può superare la mera unità frastica, cosa che rende inattendibile e poco significativo questo criterio.

Anche Anscombe (2000: 12-13) ritiene che la nozione di brevità sia troppo vaga e parte da questa considerazione per giungere a introdurre tre paradigmi chiave per il proverbio: essere testi⁷ chiusi, autonomi e minimi. Per testi chiusi, egli intende la capacità del proverbio di essere oggetto di un'enunciazione autosufficiente, ossia che non richieda enunciazioni precedenti o successive per formare un discorso completo. Parlando di testi autonomi, egli mira a sottolineare il fatto che il proverbio può essere inserito in qualsiasi posizione del testo, senza ricorrere in un posto fisso. Per testi minimi, egli intende l'impossibilità di reperire in uno stesso proverbio un altro testo chiuso, autonomo e minimo che sia, a sua volta, un proverbio.

Riteniamo che parlare di enunciazione autosufficiente sia quanto di più preciso e corretto in relazione ai proverbi ci possa essere. Anzitutto, questa proprietà va oltre le mere categorie di brevità, concisione, ma anche di frase o testo. Allo stesso tempo, è in grado di rendere conto del perché il proverbio non possa essere un costituente di frase: in questo caso, infatti, necessiterebbe di altro per entrare nella comunicazione e non sarebbe più autosufficiente.

Ciononostante, per tutte le ragioni esposte in precedenza (cfr. § 1.1.1), crediamo che non sia necessario parlare di testi a proposito del proverbio. Quanto, però, non ci convince è un altro elemento di questa definizione. Parlando di proverbi come testi minimi, Anscombe sostiene che se un proverbio (prendendo come esempio testi più lunghi come le favole di De La Fontaine) dà vita a un altro proverbio (di norma più breve), perdendo così la caratteristica di essere un testo minimo, esso cesserà di essere un proverbio. I nostri dati mostrano, al contrario, come, almeno in una prospettiva sincronica, possano coesistere proverbi più lunghi e le relative versioni ridotte, come emerge negli esempi (4) e (5), nonché proverbi originati dalla modifica di altri proverbi e le relative versioni "originali". Si mostrerà più avanti, ad esempio, che saranno i proverbi modificati a non essere più, in occasioni puntuali, proverbi (per il fenomeno di *déproverbialisation* e *détournement*) e non i proverbi che hanno dato vita a questi ultimi a cessare di essere tali. Il fatto che una frase non venga più riconosciuta come proverbio è dovuto all'uso che ne fanno i parlanti. Se un proverbio non viene più utilizzato o

⁷ In realtà, egli parla di volta in volta di *discours* e di *textes*, ma noi preferiamo utilizzare il termine testo intendendo sia il testo scritto sia il discorso orale.

si afferma in una variante diversa, i parlanti che non hanno più memoria del proverbio o della variante originale non riconosceranno più questo proverbio come tale ma rimarrà pur sempre un proverbio, sebbene in disuso.

Concentrandosi sulla caratteristica di chiusura e autonomia dei proverbi, si scorgerà qualcosa di simile in Schapira (2000), che ritiene che i proverbi siano frasi generiche che veicolano un messaggio compiuto, nonché autonomo dal punto di vista referenziale e grammaticale. Questo comporta ulteriori vincoli linguistici, ossia che la frase sia un enunciato anaforico chiuso, senza rapporto deittico con l'enunciatore o con la situazione discorsiva (Schapira, 2000: 85). Su queste considerazioni, intimamente legate alla genericità dei proverbi, torneremo in seguito. Per il momento, ci focalizzeremo sulla dimensione fraseologica del proverbio e sulla sua autonomia.

Analizziamo le nozioni di autonomia e chiusura del proverbio come definite da Anscombe (2000: 12). Il proverbio sarebbe in grado di formare un enunciato autosufficiente, tale da non richiedere ulteriori enunciazioni anteriori o posteriori, né tantomeno formule che lo introducano o ne mostrino lo status paremiologico⁸. Se consideriamo il proverbio come frase, o come enunciato, ci rendiamo conto di quanto tali precisazioni siano superflue, in quanto racchiuse per definizione nella nozione di frase e di enunciato. Parlando di enunciato, e non di testo, possiamo sostenere che non sia necessaria alcuna enunciazione anteriore o posteriore, né tantomeno che sia richiesto un qualche elemento per completarla, dato che di per sé un enunciato è già un'espressione satura. In ragione di ciò, l'enunciato non deve per forza mantenere rapporti grammaticali o sintattici con il cotesto. Ecco che la nozione di autonomia è già racchiusa nella nozione di enunciato.

Siamo convinti che l'autore giunga a queste considerazioni perché utilizza come punto di partenza il confronto tra proverbi ed espressioni idiomatiche, le quali hanno certamente bisogno di essere completate e di instaurare relazioni grammaticali, sintattiche e referenziali con il cotesto e il contesto. Tuttavia, confrontare proverbi ed espressioni idiomatiche può essere, al contempo, estremamente utile ma anche forviante, poiché le grandezze comparate sono incommensurabili: da un lato, come si è già dimostrato, ci troviamo di fronte a frasi, dall'altro siamo in presenza di costituenti di frasi, assimilabili a nomi, predicati verbali, aggettivi, avverbi, ecc.

⁸ Sulle formule introduttorie e sulla loro omissione torneremo successivamente (cfr. § 4).

Per contro, l'autonomia del proverbio così come concepita da Shapira (2000), ossia da un punto di vista meramente referenziale, riesce a definire una condizione necessaria del proverbio, in grado di dare conto a sua volta della differenza tra il proverbio e gli altri tipi di frasi ed enunciati. Di questo si parlerà più dettagliatamente nel § 1.4.3, quando si tratterà del rapporto tra proverbi e denominazione, nonché nel § 1.5, quando si esporranno i casi di *déproverbialisation* e *détournement*.

Ora focalizziamoci sulle altre condizioni necessarie del proverbio, che possiamo reperire mediante il raffronto tra proverbi e altri fenomeni fraseologici e sentenziosi, cominciando proprio dal confronto tra proverbi e espressioni idiomatiche.

1.1.3. Espressioni idiomatiche

Gli studi sulle espressioni idiomatiche vanno di pari passo con le ricerche sulla fissità, un fenomeno che abbraccia una categoria molto ampia di espressioni, che va dalle espressioni idiomatiche propriamente dette a qualsiasi tipo di espressione polilessicale. Su questo tema si innesta quello della composizionalità semantica. Si è, così, deciso di cominciare questa breve sezione dedicata alle espressioni idiomatiche con una piccola trattazione della fissità e della composizionalità.

Per prima cosa, interrogiamoci su che cosa sia la fissità. In maniera generale, la fissità fa parte della lingua ed è una *conditio sine qua non* del suo funzionamento sociale. Senza fissità non potrebbe esistere l'associazione indissolubile significato-significante. Il segno linguistico, per essere utilizzato e riconosciuto correttamente, deve essere fisso. Si tratta di ciò che Saussure chiama *l'immutabilité du signe*. Se i segni non fossero stabili e se ogni individuo avesse la libertà di creare nuovi segni linguistici a suo piacere, non vi sarebbe più funzionamento sociale possibile per la lingua (Polguère, 2008: 36-37).

Oltre al segno linguistico, la fissità può interessare altresì le sequenze polilessicali (Gross, 1996: 9). Il cosiddetto fenomeno della polilessicalità si manifesta quando si è in presenza di sequenze formate da almeno due parole aventi un'esistenza propria e autonoma in lingua. Si escluderanno, dunque, da questo ambito, i fenomeni di derivazione per suffissazione o prefissazione, dato che suffissi o prefissi, salvo qualche eccezione riconducibili a parole derivanti da lingue antiche come il greco o il latino, non sono parole autonome che vengono cioè utilizzate da sole.

Il tema della fissità, oltre ad essere di grande rilevanza dal punto di vista sintattico, è strettamente legato alla semantica delle espressioni polilessicali (cfr. Gross, 1996: 10-11): tali

espressioni funzionano in blocco e da questo risulta che il senso “ordinario” dell’espressione, ossia il senso compositivo, che deriva dalla somma dei significati dei suoi componenti, non coincide con il senso idiomatico della sequenza. In realtà, Gross (1996: 10) precisa che affinché una sequenza polilessicale sia fissa, questa deve essere semanticamente opaca. Ora, nel presente studio, preferiamo non parlare di opacità e trasparenza. Come mostrato da Casadei (1996: 391), la relazione tra significato letterale e significato idiomatico non è necessariamente opaca. Ben l’85,6% delle espressioni idiomatiche analizzate da Casadei ha un significato idiomatico motivato dal significato letterale, il che indica che una netta maggioranza di espressioni idiomatiche è trasparente. Anche Gibbs (1994: 290-295) dedica ampio spazio alla motivazione metaforica del significato idiomatico delle espressioni idiomatiche, in opposizione a quanti sostengono che il significato delle frasi idiomatiche sia determinato arbitrariamente. Dimostra nella fattispecie quanto un’espressione idiomatica come *Spill the beans* evochi, in un campione di individui interrogati separatamente, le stesse immagini mentali. Per tutte queste ragioni, preferiamo parlare di *composizionalità* e di *non composizionalità*.

Spieghiamo, dunque, ciò che si intende per significato compositivo e non compositivo. Un’espressione complessa è compositiva se il suo significato complesso conserva i significati delle parole che la compongono. Come esemplifica Prandi (2006: 73), “il significato di un’espressione come *avere una casa in montagna*, per esempio, conserva i significati di *avere*, di *casa* e di *montagna*”. Al contrario, un’espressione complessa è non compositiva se il suo significato è indipendente dal significato delle parole che la compongono. “L’espressione *avere il vento in poppa*, per esempio, significa più o meno «essere in un momento fortunato». Nel suo significato non c’è traccia del significato di *vento* e di *poppa*” (Prandi, 2006: 73).

Questo è quanto accade se un locutore che ha una buona conoscenza di una lingua straniera si trova di fronte a un’espressione idiomatica e non ne riesce a riconoscere il significato complesso o idiomatico: sebbene conosca il significato delle singole parole, non giunge a interpretare correttamente la sequenza. Certo è che una stessa espressione complessa può avere occasionalmente una lettura compositiva e una lettura non compositiva. Ad esempio, l’espressione francese *Les carottes sont cuites* può avere una doppia lettura: da un lato, può indicare che la verdura in questione è cotta e pronta per essere mangiata, dall’altro potrebbe significare che la situazione è disperata e non vi è più nulla da fare (Gross, 1996: 11).

La prima interpretazione è compositiva, mentre la seconda è non compositiva o idiomatica.

Secondo Gross (1996: 12), la fissità è intrinsecamente legata alla sintassi. Come numerosi altri autori⁹, considera che il blocco delle proprietà trasformazionali sia una *conditio sine qua non* della fissità. In effetti, egli sottolinea come le frasi libere abbiano delle proprietà trasformazionali che dipendono dalla loro organizzazione interna. Questo significa che data una frase come *Il bambino ha letto il libro*, si possono applicare certi cambiamenti strutturali denominati “trasformazioni”. Esempi di trasformazioni della frase in oggetto sono: passivizzazione (*Il libro è stato letto dal bambino*); pronominalizzazione (*Il bambino l’ha letto*); dislocazione (*Il libro, il bambino l’ha letto*); frase scissa (*È il libro che il bambino ha letto*); relativizzazione (*Il libro che ha letto il bambino*); interrogazione (*Che cosa ha letto il bambino? – Il libro*); sostituzione lessicale per sinonimia (*Il ragazzino ha dato una letta al romanzo*). Se prendiamo in esame l’espressione idiomatica *Tirare le cuoia*, capiamo che non possiamo produrre enunciati come **Le cuoia sono state tirate da Tizio*; **Tizio le ha tirate (le cuoia)*; **Le cuoia le ha tirate Tizio*; **Sono le cuoia che ha tirato Tizio*; **Che cosa ha tirato Tizio? – Le cuoia*; **Tizio ha lanciato il corame*. Gli esempi appena prodotti sono agrammaticali o per lo meno non producono enunciati che mantengono lo stesso significato idiomatico dell’espressione *Tirare le cuoia*. Spesso, infatti, le trasformazioni sono possibili ma implicano la perdita del significato non compositiva, essendo realizzabili solo con frasi libere. Se pensiamo a espressioni idiomatiche che permettono una doppia interpretazione (tanto compositiva quanto non compositiva), come *Les carottes sont cuites* o *Alzare il gomito* o *Vuotare il sacco*, non si può escludere che esse siano oggetto di diverse tipologie di trasformazioni sintattiche, ma in questi casi perderanno la loro natura idiomatica. Infine, non si può certo escludere che alcune espressioni idiomatiche possano essere oggetto di trasformazioni e modifiche senza per questo perdere il loro carattere idiomatico. A titolo di esempio, si può dire sia *L’alunno è stato preso in giro da tutta la classe*, che non vuole dire altro che *Tutta la classe ha preso in giro l’alunno*, cosa che dimostra come a volte il significato idiomatico possa rimanere anche in presenza di trasformazioni sintattiche quali la passivizzazione. Ecco perché è necessario tenere presente che non esiste un solo tipo di fissità, bensì diversi gradi di fissità che caratterizzano le sequenze polilessicali fisse. In realtà, non sarebbe errato ipotizzare un vero e proprio continuum tra le sequenze completamente libere e

⁹ Ad esempio, cfr. Anscombe (2005), Casadei (1996) e Vietri (1985).

le sequenze completamente fisse, cosa che emerge applicando i test sintattici proposti qui sopra e notando che la loro validità non si può estendere a priori a tutte le sequenze polilessicali fisse. Come ricorda Mejri (2005, 183-187), se il fenomeno della fissità è estremamente frequente in lingua (in media, il 20% dei testi sarebbe composto da segmenti fissi), la fissità assoluta, ossia il grado massimo di fossilizzazione di una forma sintagmatica, riguarda mediamente un'espressione su dieci.

Inoltre, le proprietà trasformazionali non riguardano solamente le espressioni idiomatiche aventi almeno un verbo, ma anche i sostantivi. Se pensiamo a una sequenza nominale fissa, composta da un sostantivo e un aggettivo, come *Anima gemella*, notiamo che non possiamo applicare quelle modifiche che sarebbero altrimenti possibili con sequenze simili ma non fisse. Non possiamo dire: **Un'anima molto gemella*; **Un'anima particolarmente gemella*; **La gemella di ques'anima*; **Quest'anima è gemella*. Per contro, le stesse trasformazioni sarebbero possibili con una sequenza libera come *Un libro difficile*. Potremmo, infatti, dire: *Un libro molto difficile*; *Un libro particolarmente difficile*; *La difficoltà di questo libro*; *Questo libro è difficile*. Come nel caso delle espressioni idiomatiche, la fissità parziale riguarda altresì i sintagmi nominali e i nomi composti. Gross (1988), infatti, dimostra che i composti assolutamente fissi sono solamente il 10% dei gruppi nominali che presentano fenomeni di fissità, seppure di diverso grado o entità.

Come si sarà inteso, la fissità non interessa unicamente le espressioni idiomatiche, ma anche le locuzioni, che sono, appunto, gruppi di elementi che non possono essere attualizzati singolarmente (Gross, 1996: 14). Anche l'attualizzazione, o meglio la mancanza di attualizzazione degli elementi di una sequenza fissa, è indice di fissità. Pertanto, anche se nel presente studio si darà molta rilevanza alle espressioni idiomatiche *tout court*, è bene ricordare che la fissità è un fenomeno che può riguardare altresì sintagmi verbali, nominali, aggettivali, avverbiali oppure preposizionali.

1.2. Proverbi ed espressioni idiomatiche: un confronto sintattico

Si è già visto e dimostrato che i proverbi sono frasi. Le espressioni idiomatiche, dal canto loro, sono costituenti di frasi. Non si è fatta distinzione, nel paragrafo precedente, così come avviene in letteratura, tra espressioni idiomatiche e locuzioni, considerato il loro comportamento sintattico simile. In effetti, sia le espressioni idiomatiche sia le locuzioni condividono la caratteristica di essere unità inferiori alla frase, ovvero costituenti di frasi. Tale proprietà contrappone le espressioni idiomatiche ai proverbi.

La nostra impressione è che non si insista abbastanza su questa piccola differenza che è, tuttavia, in grado di guidare il ricercatore nel confronto tra proverbi ed espressioni idiomatiche. Per questa ragione, non crediamo che sia azzeccata la scelta di presentare come esempio di espressione idiomatica *Les carottes sont cuites*, che invece è un vero e proprio esempio da manuale presentato in numerosi studi, in ragione del fatto che si tratta di una frase a tutti gli effetti e non di un costituente come potrebbe essere *To kick the bucket*, altro classico esempio da manuale, questa volta in ambito anglosassone. Almeno dal punto di vista sintattico, *Les carottes sont cuites* assomiglia molto di più a un proverbio che a un'espressione idiomatica, sebbene non funzioni semanticamente come un proverbio, in quanto manca di genericità (cfr. § 1.4.2 e 1.4.3), e possa essere definito, in linea con quanto asserito da Anscombe (2000: 10), *frase situazionale*.¹⁰

Tra gli studi relativi al rapporto tra proverbi, espressioni idiomatiche e fissità, sono degni di nota i lavori di Jean-Claude Anscombe (2005). Egli si è servito del raffronto tra proverbi ed espressioni idiomatiche per criticare e smentire la tesi della fissità dei proverbi fornendo le seguenti motivazioni.

In primis, egli sostiene che poiché esiste un substrato comune a proverbi di aree geografiche molto distanti, che permette di avere proverbi simili (se non identici) in lingue diverse, e dato che tale base comune non esiste per le espressioni idiomatiche, il proverbio non può essere considerato fisso come le espressioni idiomatiche. L'autore si riferisce, chiaramente, alle lingue parlate in Europa e specialmente alle lingue indo-europee, che presentano un ampio terreno culturale condiviso grazie all'influenza del latino. Pensiamo, ad esempio, al proverbio *Una hirundo non facit ver* che è presente in molte lingue europee, come in italiano, *Una rondine non fa primavera*, in francese, *Une hirondelle ne fait pas le printemps*, in spagnolo, *Una golondrina no hace verano*¹¹, in portoghese, *Uma andorinha não faz primavera*, in inglese, *One swallow does not make a summer*, in tedesco, *Eine Schwalbe macht keinen Sommer*,¹² solo per citare alcune lingue. Notiamo che, nonostante la fortissima

¹⁰ Anscombe (2000, 10) nota che frasi come *Il aura passé de l'eau sous les ponts; C'est la goutte d'eau qui fait déborder le vase; Quand il n'y en a plus, il y en a encore* sono frasi episodiche e non generiche, come invece i proverbi (cfr. § 2.6.2 e 2.6.3). Anscombe separa frasi di questo tipo dai proverbi, in ragione della loro mancanza di genericità, e le definisce frasi situazionali.

¹¹ Il quesito è: perché in spagnolo si dice "verano", ossia estate, sebbene il clima della Spagna sia molto simile a quello dell'Italia? Molto probabilmente nel passaggio dal latino allo spagnolo si mantenne la radice etimologica *ver* da cui, in effetti, deriva la parola spagnola *verano*, cosa che invece non è successa nella versione usata dal limitrofo popolo portoghese che, al contrario, parla di primavera.

¹² Gli esempi in francese, spagnolo, portoghese, inglese e tedesco sono tratti da Anscombe (2005: 22). Notiamo che anche uscendo dallo spazio delle lingue indo-europee, pur rimanendo in Europa, possiamo trovare gli stessi

somiglianza tra questi proverbi, permangono differenze a nostro avviso non trascurabili, probabilmente legate alle diverse realtà climatiche caratterizzanti le aree in cui si sono sviluppate le lingue sopraccitate. In italiano, francese, portoghese, così come in latino, si prende come riferimento temporale la *primavera*, mentre in spagnolo¹³, inglese e tedesco la stagione prescelta è l'*estate*. Tali differenze rendono unici i proverbi in ogni lingua e devono essere osservate e rispettate da un professionista delle lingue per produrre proverbi effettivamente esistenti che non diano adito ad interpretazioni errate, specialmente dal punto di vista pragmatico (ad esempio, l'enunciato *Una rondine non fa estate* provocherebbe in un nativo italiano una serie di inferenze finalizzate a giustificare tale slittamento lessicale). Pertanto, è chiaro che, sebbene esistano proverbi *simili* in ragione della loro origine comune, questi sono adottati dalle culture di arrivo e assumono i tratti distintivi delle stesse. È certamente innegabile che vi sia un ampio *fonds commun*, come viene definito da Anscombe (2005: 23), alla base di numerosi proverbi, sebbene il tutto debba essere ristretto all'area europea (o, tutt'al più, alle culture in cui si parlano lingue nate in Europa), condizione a nostro avviso non trascurabile.

Contro questa tesi, si possono avanzare almeno due contro argomentazioni. In primo luogo, è facile individuare, accanto a tale substrato comune, proverbi che non hanno la stessa origine e che quindi hanno significati compositivi molto diversi. Pensiamo, a mero titolo di esempio, al proverbio *Si parla del diavolo e spuntano le corna*. In francese scompare completamente il riferimento al binomio *diavolo e corna*, che viene sostituito dal binomio *lupo e coda*: *Quand on parle du loup on en voit la queue*. In spagnolo, la distanza aumenta, dato che si fa riferimento all'ambito della religione o, meglio, al rapporto tra religione e potere temporale del papa, sostenendo che *Hablando del rey de Roma por la puerta asoma*. Sono, questi, proverbi molto frequenti nelle tre lingue, ma che fanno riferimento a circostanze, campi semantici e referenti che nulla hanno in comune, se non il fatto di essere personaggi malvagi (in italiano abbiamo il demonio, in francese un animale tradizionalmente feroce e in spagnolo il papa, visto però nella sua veste di detentore di un potere temporale, molto probabilmente legato alla vicenda storica dello spostamento della corte papale da Roma ad

proverbi; pensiamo, a titolo di esempio, alla lingua basca, '*Enara batek ez du udaberrria egiten*' (trad. lett.: 'Una rondine non fa primavera') e alla lingua ungherese, '*Egy fecske nem csinál nyarat*' (trad. lett.: 'Una rondine non fa estate'). La presenza di questo proverbio in lingue non appartenenti alla famiglia delle lingue indo-europee, ma comunque in stretta relazione con le culture europee, mostra come tali proverbi non siano presenti nelle diverse culture per un mero fattore etimologico, bensì per i contatti culturali e commerciali tra popoli diversi.

¹³ Si ritiene, come già spiegato in nota 12, che in spagnolo la ragione dell'uso della parola "estate" sia più imputabile a un discorso etimologico che alla realtà climatica locale.

Avignone¹⁴, dato che in origine si usava la parola *ruin*, ossia ‘vile, spregevole, meschino, avaro’ al posto di *rey*, che metteva in luce caratteristiche molto vicine a quelle del demonio). Non si può, dunque, escludere che i tre proverbi abbiano attinto alla stessa combinazione di concetti (un *parlare* di un *essere spregevole* di cui ad un tratto *si intravede un elemento costitutivo*), sebbene non sia immediato individuare tali parallelismi e benché non sia necessario conoscere la storia di un proverbio perché sia utilizzato correttamente (si veda più avanti, in questo §, l’esempio di *Tutti i nodi vengono al pettine*, in cui il *pettine* in questione ha subito uno slittamento semantico dalla fine della società preindustriale ad oggi, ma pur non conoscendo queste ragioni filologiche, un parlante medio riesce ad usare il proverbio correttamente). Se si analizzano proverbi maggiormente legati a certe realtà locali, si nota che la somiglianza tra proverbi di lingue diverse non trova più giustificazione. Pensiamo al proverbio italiano *Al contadino non far sapere quanto è buono il formaggio con le pere*. Sulla base delle nostre ricerche, non risulta che esistano proverbi uguali o simili in francese, spagnolo, inglese e portoghese. Ci siamo inoltre avvalsi della monografia di Massimo Montanari, uscita nel 2008 con il titolo *Il formaggio con le pere: la storia in un proverbio*, che indaga le origini e il significato di questo proverbio, basandosi su numerosi documenti europei dal medioevo al XIX secolo relativi all’arte culinaria, alla medicina, alle credenze popolari, agli usi e ai costumi, oltre che dizionari e raccolte di proverbi. Si può così constatare che il tema del formaggio e delle pere era presente anche in altre culture sia nel Medioevo sia nella prima Età moderna: da questo connubio consolidato trae origine il proverbio francese nato nel Medioevo *Entre le fromage et la poire chacun dit sa chanson à boire*, che nasce dall’abitudine di mangiare il formaggio e le pere a fine pasto, momento in cui i commensali sono più allegri e, dunque, più propensi a parlare e a scherzare. Questo proverbio ha dato origine all’espressione ancora oggi in uso in Francia, *Entre la poire et le fromage*, con cui si indica la fine del pasto, quando, appunto, la conversazione si fa meno seria (Montanari, 2008: 15). La pera e il formaggio erano dunque presenti nel repertorio paremiologico francese, così come in quello spagnolo, solo per citare due lingue e culture particolarmente affini a quella italiana; tuttavia, in queste due lingue, non vi è traccia di un proverbio simile a quello italiano. Il substrato culturale comune esiste, il formaggio e le pere hanno presso le culture di cui sopra

¹⁴ Si veda per l’origine di questo proverbio <http://blogs.20minutos.es/yaestaellistoquetodolosabe/cual-es-el-origen-de-la-expresion-hablando-del-rey-de-roma/>, ultima consulta il 15 febbraio 2014.

lo stesso valore (connubio di sapori ed alimenti particolarmente indicato a fine pasto), ma ciò non basta a ricreare o riproporre lo stesso proverbio nelle diverse culture.

Se si prendono in considerazione proverbi in uso presso culture molto distanti, questo *fonds commun* non avrà più ragione di esistere. In letteratura paremiologica si riportano spesso esempi di proverbi asiatici, ma anche africani, ebraici, rumeni e di altre svariate provenienze. Ebbene, tali proverbi non hanno nulla in comune con il repertorio paremiologico francese o italiano, non condividendo nemmeno i valori trasmessi. Se pensiamo a proverbi antichi, riportati da testi antichi come la Bibbia, ma anche da Aristotele o dagli antichi romani, notiamo con stupore una serie di proverbi che veicolano valori diversi dai nostri, che sconfinano, ai nostri occhi, nell'immoralità, nella crudeltà umana o nella vendetta (Shapira, 1999: 90-91). Ciononostante, siamo portati a interpretare correttamente anche proverbi mai sentiti prima appartenenti a culture lontane nello spazio o nel tempo (Gibbs, 1994: 309-317). Il fatto che si possano comprendere questi proverbi, sebbene questi ci possano sembrare distanti o alquanto bizzarri, è spiegabile in ragione del particolare funzionamento semantico dei proverbi, di cui si tratterà in dettaglio più avanti quando si parlerà della loro genericità e della loro capacità di essere denominazioni (cfr. § 1.4.2 e 1.4.3). Ecco che si rivela molto più interessante, per capire meglio il funzionamento semantico dei proverbi, interrogarsi sul perché riusciamo a decifrare e interpretare correttamente proverbi asiatici o africani, piuttosto che cercare somiglianze tra proverbi di culture affini, che hanno avuto storicamente una base culturale comune da cui hanno senz'altro ereditato una serie di *doxa* e *stereotipi* (per maggiori approfondimenti, cfr. Schapira, 1999: 15-45) che hanno a loro volta dato vita a elementi linguistici della stessa natura.

Si è, pertanto, tentato di dimostrare come la tesi della non fissità dei proverbi, che poggia sulla somiglianza di proverbi presenti in culture diverse, si basi su argomentazioni discutibili. Anscombe (2005: 23) sostiene, inoltre, che le frasi situazionali (cfr. nota 10) godono di un alto grado di fissità perché, come le espressioni idiomatiche, non attingono da questo terreno comune. Anche in questo caso, non si può parlare di una realtà assoluta, in quanto la frase situazionale francese che cita come esempio (Anscombe, 2005: 23), ossia *Il a passé de l'eau sous les ponts*, in opposizione all'altra frase situazionale spagnola, *Ha llovido (desde entonces)*, per mostrare che le due sono formalmente diverse¹⁵, in realtà è l'esatta

¹⁵ Dimostrando che le due frasi situazionali hanno diversi sensi letterali nelle due lingue, l'autore giunge alla conclusione che esse sono più simili alle espressioni idiomatiche che ai proverbi, ribadendo il fatto che i proverbi provengano da un fondo comune e le espressioni idiomatiche no. L'autore stesso, tuttavia, puntualizza che

traduzione della stessa frase situazionale italiana, che ha pari frequenza d'uso e stesso significato: *Il a passé de l'eau sous les ponts* e *Ne è passata di acqua sotto i ponti*. Si pensi, inoltre, a quelle espressioni idiomatiche che sono molto simili in lingue diverse, come *Dare una mano*, *Mettere in luce*, *Buttare/dare un occhio*, che si possono tradurre, rispettivamente, in inglese con *To give a hand*, in francese con *Mettre en lumière*, in spagnolo con *Echar un ojo*. E ancora pensiamo al nostro *Cercare un ago in un pagliaio* che ritroviamo molto simile in francese con *Chercher une aiguille dans une botte de foin*. L'immagine evocata è la stessa, così come il concetto espresso, ossia l'enorme difficoltà nel reperire un elemento all'interno di un insieme di elementi smisuratamente più grande e caotico. Ecco dunque l'ennesima dimostrazione di quanto queste argomentazioni portate contro la tesi della fissità dei proverbi siano troppo legate alla realtà di ogni singolo proverbio, di ogni espressione idiomatica e di ogni combinazione linguistica e non della categoria generale dei proverbi o della categoria generale delle espressioni idiomatiche.

La seconda prova addotta da Anscombe a sostegno della sua tesi sulla non fissità dei proverbi riguarda le varianti dei proverbi. Questo è lampante: tutti i proverbi hanno (o possono avere) delle varianti che sono utilizzate come e quanto le "versioni originali", per così dire, repertorate dai dizionari di proverbi. Secondo l'autore, questa libertà non è possibile con le espressioni idiomatiche, se non in misura molto ridotta e del tutto eccezionale. Lo stesso Anscombe (2005: 24) ammette, dunque, che le espressioni idiomatiche possono variare, benché meno frequentemente. In realtà, sebbene non abbiamo trovato statistiche relative alla variabilità delle espressioni idiomatiche, siamo convinti che si tratti di un fenomeno di grande rilevanza. Pensiamo, ad esempio, ai casi in cui il verbo di un'espressione idiomatica non sia fisso. Numerosi esempi di questo tipo sono forniti da Casadei (1996: 94-95): *Ammainare / calare le reti*; *Andare / navigare con il vento in poppa*; *Fare / costruire / fabbricare castelli in aria*; *Essere / stare / andare alla deriva*. Altri esempi che abbiamo reperito nella vita di tutti i giorni sono: *To give / to get the green light*; *Trovare / cercare il pelo nell'uovo*. Vi sono poi casi in cui le varianti di espressioni idiomatiche sono date dalla presenza di sinonimi a livello dei sostantivi presenti nelle espressioni, come in *Carlarsi i pantaloni / le mutande*; *Prendere una brutta / cattiva piega / china*; *Condurre la nave / la barca in porto* (esempi tratti da Casadei, 1996: 94).

questo non è sempre valido, esistendo espressioni idiomatiche in lingue diverse che, invece, coincidono (Anscombe, 2005: 23).

Tra le argomentazioni principali relative alla fissità delle espressioni idiomatiche e alla non fissità dei proverbi, troviamo altresì una spiegazione lessicale, che verrà ripresa anche per affermare che il significato dei proverbi è compositivo. Il fatto che i proverbi si siano attualizzati nei secoli è la dimostrazione del fatto che essi siano stati impiegati e vengano tutt'ora impiegati non tanto in virtù del loro significato inteso, idiomatico, non compositivo, quanto piuttosto in ragione del loro senso letterale e dunque compositivo, vale a dire come una qualsiasi sequenza polilessicale libera. Anscombe (2005: 25) porta numerosi esempi tratti dal francese, il più rappresentativo dei quali è il seguente. Se nel XIV-XV secolo il proverbio attestato era *En forgeant devient len fevre*, nel XVI secolo si ritrova lo stesso proverbio già modificato in *En forgeant on devient orfevre*, per poi giungere alla versione del proverbio che risulterà ben più comune ai francofoni e francesisti di oggi, ossia *C'est en forgeant qu'on devient forgeron*.

La prima osservazione che possiamo avanzare sulla base di quanto già presente in letteratura, è che anche le espressioni idiomatiche possono subire delle attualizzazioni lessicali. Balibar-Mrabti (2005: 48) mostra un esempio in cui l'espressione idiomatica *Brûler ses vaisseaux* è presente in un romanzo¹⁶ recente secondo la variante attualizzata *Brûler ses bateaux*, prova di quanto anche le espressioni idiomatiche possano subire un processo di attualizzazione.

In secondo luogo, non possiamo affermare che i proverbi vengano sempre attualizzati: diversi di questi, infatti, mantengono ancora oggi termini arcaici (Schapira, 2000: 90). Per fare un esempio, pensiamo al proverbio italiano *Il troppo stroppia*. *Stroppiare* è una variante di *storpiare* e, per quanto risulta da qualche ricerca su *webcorpora*, non viene oggi utilizzato se non nel proverbio appena citato (nella stragrande maggioranza dei casi) o comunque come verbo se stante, ma non tanto nel suo significato originale di *storpiare* (benché si trovi anche con questa accezione come variante regionale di *storpiare*), bensì con il significato che assume agli orecchi dei parlanti di oggi e che si vede inevitabilmente influenzato dal proverbio stesso, ossia *eccedere; essere troppo; essere smisurato*. Si veda il seguente esempio: “A volte stroppiare un po' è consentito. #golden #retrò #occhiali montatura #oro e #tartaruga”¹⁷. E ancora, pensiamo al proverbio: *Tutti i nodi vengono al pettine*. Il pettine a cui

¹⁶ Nothomb, A. (2005). *Antéchrista*. Paris: Le livre de poche, p. 132.

¹⁷ Dal forum rintracciabile al seguente URL: <http://depop.com/it/chiarasunny/a-volte-stroppiare-un-po>, ultima visita effettuata il 24 febbraio 2014. Il forum riguarda accessori e altri oggetti di moda. Nel caso specifico, si

oggi pensiamo quando citiamo o sentiamo questo proverbio è il comune pettine usato per pettinarsi, mentre in origine si riferiva al pettine del telaio, “che non fa passare i nodi della matassa di lana o di cotone, quando questa non è stata preliminarmente scardata” (Bronzini, 1999: 50-51). Etimologicamente e storicamente, il nome *pekten* si è diffuso in area greco-latina ed è stato applicato al pettine da telaio, strumento su cui il tessitore regola il filato, ovvero fa passare uno o più fili del tessuto da tessere in ogni dente. Pertanto, originariamente il riferimento al tipo di pettine e al materiale che vi doveva passare attraverso era chiaro. Successivamente, con il passaggio da una società agricola a una industriale, si ha avuto la perdita progressiva dei telai (e quindi di *quei pettini*) dalle campagne, cosa che ha provocato la sostituzione quasi naturale del referente perduto con un referente invece più comune. Come questo proverbio ne esistono sicuramente altri in cui una parola persiste ma segue gli sviluppi della società e cambia riferimento e accezione (Bronzini, 1999: 51-53). La riflessione è che quand’anche le parole dei proverbi ci possano sembrare attuali, in realtà potrebbero essere rimaste quelle originali. Sono stati l’uso e i cambiamenti della società ad aver fatto assumere interpretazioni errate (ovvero attuali e attualizzate) a queste parole. Pertanto, anche i proverbi, sebbene in misura probabilmente minore rispetto alle espressioni idiomatiche, hanno conservato caratteri lessicali antichi e a volte desueti, senza essere per questo attualizzati.

Un’altra argomentazione utilizzata da Anscombe per confutare la tesi della fissità dei proverbi è la ripresa anaforica, impossibile per le sequenze fisse come le espressioni idiomatiche e, al contrario, frequente per i proverbi (Anscombe, 2005: 25). Crediamo che questo comportamento dissimmetrico possa essere spiegato con la diversità di gradi di fissità di espressioni idiomatiche e proverbi: è fuori questione che le prime siano più fisse rispetto ai proverbi, i quali sono sicuramente meno rigidi, probabilmente in virtù del fatto che sono frasi e non semplicemente costituenti di frasi come le espressioni idiomatiche. Ad ogni modo, torneremo più avanti anche su questa questione.

1.3. Proverbi ed espressioni idiomatiche: un confronto semantico

Secondo molti, i proverbi e le espressioni idiomatiche non funzionano in maniera simile dal punto di vista semantico. Crediamo che le osservazioni di Gaston Gross (1996) riescano a sintetizzare bene le motivazioni generalmente addotte per considerare che la semantica dei proverbi sia molto diversa da quelle delle espressioni idiomatiche.

mostra, in una fotografia legata al commento sopra riportato, un paio di occhiali alquanto appariscenti. In base al contesto, il verbo stroppiare assumerebbe il significato di “eccedere nella stravaganza”.

Si parte dall'idea che i proverbi siano e funzionino come le poesie, le canzoni e le preghiere. Questi testi condividono la caratteristica di essere fissi dal punto di vista discorsivo. Gross (1996: 143-144) parla, a tal proposito, di "*figement discursif*", che è un tipo di fissità che impedisce qualsiasi cambiamento formale in queste tipologie testuali. Di solito, sono i testi ad uso collettivo a presentare questo tipo di fissità. Si tratta di prodotti preconfezionati presi tali quali e utilizzati per determinati usi sociali. Essendo parte integrante della cultura di un popolo, essi sono noti ai parlanti appartenenti alla cultura presso cui si sviluppano, ma data la loro fissità e il loro status citazionale devono essere imparati a memoria.

L'unico tratto che hanno in comune con le sequenze fisse è precisamente questo: il fatto di non essere il frutto della creazione individuale e di essere, dunque, elementi stabili a cui non si può apportare alcun cambiamento. Tuttavia, le sequenze fisse, secondo Gross (1996), vanno oltre la mera fissità intesa nei termini di non possibilità di cambiamento (salvo a fini ludici). Le sequenze fisse hanno un tipo di fissità che è linguistico, legato al blocco delle proprietà trasformazionali e alla non attualizzazione di cui si è parlato in precedenza.

Tali considerazioni portano a pensare che i proverbi siano fissi dal punto di vista sociale e testuale ma non dal punto di vista linguistico. Da questa piccola affermazione scaturiscono importanti conseguenze, specialmente dal punto di vista semantico: le frasi proverbiali "*ne sont nullement figées linguistiquement, puisqu'elles ont une syntaxe régulière (compte non tenu de l'archaïsme qui caractérise les proverbes anciens) et que leur sens est compositionnel, c'est-à-dire transparent*" (Gross, 1996: 144). Non siamo certi che questa affermazione sia valida, specialmente perché i dati in nostro possesso parlano chiaro, mostrando che i proverbi si interpretano difficilmente in maniera compositazionale e quasi sempre in maniera idiomatica o non compositazionale.

Siamo tuttavia di fronte a un problema di grandissima rilevanza per lo studio linguistico dei proverbi, che non può essere risolto in maniera troppo sbrigativa. La semantica del proverbio è compositazionale o non compositazionale? Il proverbio funziona similmente o diversamente dall'espressione idiomatica? Tutti questi interrogativi comporteranno notevoli conseguenze anche per la traduzione o la trasposizione del proverbio in L2.

1.3.1. Una risposta: la vera differenza tra proverbi ed espressioni idiomatiche

Come si è visto, la trattazione delle differenze e delle somiglianze tra proverbi ed espressioni idiomatiche è molto articolata e spesso corre sul filo del rasoio delle piccole prove

linguistiche avanzate via via su ogni aspetto, linguistico e culturale, di proverbi ed espressioni idiomatiche. L'intuizione è che dal punto di vista del funzionamento semantico non vi siano in fondo tante differenze tra espressioni idiomatiche come *Cercare il pelo nell'uovo*, *Cercare un ago in un pagliaio*, *Piangere sul latte versato* o *Passare il Rubicone* e proverbi come *Tra il dire e il fare c'è di mezzo il mare*, *L'erba del vicino è sempre la più verde*, *Il lupo perde il pelo ma non il vizio* o *A caval donato non si guarda in bocca*. Come le espressioni idiomatiche citate non sono utilizzate solitamente per fare riferimento a uova, aghi, latte o fiumi, così anche i proverbi, presi per il loro significato idiomatico, non si riferiscono soltanto a un mare, un prato, un lupo o un cavallo. Ciò non toglie che è possibile usare sia le espressioni idiomatiche sia i proverbi in funzione del loro significato compositivo. In altre parole, potremmo veramente cercare un pelo sulla superficie di un uovo, cercare un ago all'interno di un pagliaio, considerare che il prato del nostro vicino di casa sia più bello o che non si debbano guardare i denti del cavallo che è stato regalato. Pertanto, come le espressioni idiomatiche possono essere oggetto di una doppia lettura o una doppia interpretazione (una compositiva e l'altra non compositiva), anche i proverbi possono avere di volta in volta una doppia interpretazione, come indicato dalla nostra esperienza di parlanti e utilizzatori di proverbi.

Vediamo che emerge poco a poco un elemento di grande importanza che non possiamo trascurare se studiamo le espressioni idiomatiche e i proverbi: il contesto d'uso.

Se analizziamo questi elementi inseriti nel loro contesto d'uso, capiamo immediatamente la principale caratteristica che potrebbe essere in grado di giustificare le differenze che sono presenti tra proverbi ed espressioni idiomatiche: i proverbi sono frasi, le espressioni idiomatiche sono costituenti di frasi.

I proverbi, dunque, non sono solo forieri di un significato (che sia esso compositivo o non compositivo poco importa). Essi veicolano, di per sé, un messaggio, essendo frasi e quindi, una volta entrati in testi o discorsi, enunciati.

Contrariamente ai proverbi, le espressioni idiomatiche da sole non possono veicolare alcun messaggio, data la loro realtà di costituenti e non di frasi. Esse necessitano di qualcos'altro, anche solo di un soggetto, per formare una frase. La frase così costruita, facendosi enunciato, veicolerà un messaggio.

Questa considerazione, apparentemente semplice e banale, è quasi sempre assente dagli studi linguistici intorno al proverbio, benché abbia, a nostro giudizio, la capacità di

guidare il ricercatore nella comprensione di quanto sia davvero funzionale per l'uso e la comprensione del proverbio.

Il fatto che il proverbio sia una frase, sebbene a volte possa anche superare la mera unità frastica, implica che difficilmente si possa confrontare con elementi che, invece, sono costituenti di frasi. O meglio, trattandosi di due grandezze incommensurabili, il loro raffronto potrà fornire solamente qualche spunto di riflessione, ma non sarà necessariamente valido.

Se consideriamo le espressioni idiomatiche alla stregua delle locuzioni verbali, nominali, avverbiali o aggettivali, ci rendiamo conto che esse non siano in grado di formare da sole un enunciato (salvo qualche rara eccezione). Al contrario, un proverbio ha la capacità di entrare nella comunicazione (ossia in testi e discorsi) anche se viene enunciato tale quale lo si trova su un qualsiasi dizionario o raccolta di proverbi. Il proverbio, è così in grado di veicolare un messaggio senza che l'enunciatore debba aggiungere formule introduttive o parti del discorso che vadano a riempire gli argomenti del proverbio, dato che questi sono, appunto, già occupati per definizione.

Per tutte queste ragioni, crediamo che si debbano rivedere alcuni parametri spesso attribuiti al proverbio. Anzitutto, non ci sembra utile, né tantomeno proficuo, distinguere tra proverbi metaforici e proverbi non metaforici, o tra *proverbi letterali*¹⁸ e *proverbi metaforici*, etichette utilizzate di frequente in numerosi studi relativi ai proverbi.

A tal proposito, si considerino i seguenti proverbi: *Tale padre, tale figlio*; *A ciascuno piace il suo*; *Anno nuovo vita nuova*; *Chi si accontenta gode*. Non è errato sostenere che in questi casi il significato pertinente sia quello compositivo. Il primo proverbio significa che il figlio è come il padre, il secondo che ogni individuo apprezza le proprie cose, il terzo che con il cambiare dell'anno si cambia anche vita e il quarto che coloro che si accontentano godono.

Si considerino ora i seguenti proverbi: *A caval donato non si guarda in bocca*; *Acqua cheta rompe i ponti*; *Batti il ferro finché è caldo*; *L'erba del vicino è sempre più verde*. Questi proverbi non vengono utilizzati per esprimere quanto intendono compositivamente, ma altro: essi indicano rispettivamente che non bisogna criticare quanto si è ottenuto in dono; che

¹⁸ Ora, precisiamo che la scelta dell'aggettivo *letterale* ci sembra infelice, dato che preferiamo utilizzare tale termine riferendoci all'interpretazione del messaggio di un enunciato. Si definisce letterale il messaggio di un enunciato quando la sua interpretazione coincide con il significato dell'espressione complessa che compone l'enunciato stesso. Viceversa, quando il messaggio non coincide con il significato dell'enunciato, avremo un'interpretazione non letterale, cosa che accade con una frequenza molto elevata nella comunicazione ordinaria (cfr. Prandi, 2004: 5-32). In questo studio, si parlerà piuttosto di *proverbi dal significato compositivo* e *proverbi metaforici*.

anche le piccole azioni possono avere grandi conseguenze; che bisogna insistere fintanto che una situazione o un tema sono aperti o di recente accadimento; che le cose che appartengono ad altri ci sembrano sempre migliori delle nostre. Si sarà notato che le parafrasi appena esposte sono estremamente diverse dai proverbi stessi. A titolo di esempio, il proverbio *A caval donato non si guarda in bocca* sarà difficilmente usato (osiamo dire, quasi mai) in presenza di un esemplare di equino regalato. Al contrario, si potrebbe supporre che il proverbio *Tale padre, tale figlio* potrebbe essere usato con un'alta probabilità in presenza di (o relativamente a) un padre e di suo figlio.

Si potrebbe, dunque, supporre che i proverbi appartenenti al primo gruppo siano proverbi dal significato compositivo, mentre quelli facenti capo al secondo gruppo siano metaforici. Tuttavia, non è sbagliato supporre che si possa enunciare un proverbio come *Tale padre, tale figlio* in situazioni in cui non siano presenti o non si faccia riferimento a un padre biologico e al relativo figlio. Tale proverbio si potrebbe applicare a personaggi o elementi che non abbiano un legame biologico, parentale o di filiazione. Ad esempio, si può enunciare questo proverbio per riferirsi a un animale e al suo cucciolo, a un artista e alla sua opera d'arte, o a un inventore e all'oggetto inventato, o ancora a un uomo politico e a un altro politico che ne segue le orme (cfr. Kleiber, 2000: 47).

Parimenti, nulla osta a impiegare i proverbi del secondo tipo, quelli cosiddetti metaforici, con significati non metaforici o compositivi. Ecco perché, a nostro avviso, si deve superare il binomio proverbio letterale, o proverbio dal significato compositivo, e proverbio metaforico.

1.4. Verso un nuovo paradigma definitorio di proverbio

Per tutto quanto è emerso fino a questo punto, crediamo che sia necessario impostare un nuovo procedimento di analisi dei proverbi, che riesca a prendere in considerazione gli aspetti davvero salienti del proverbio e che sia al contempo in grado di fornire valide spiegazioni ai diversi interrogativi posti fino ad ora.

Come si diceva, il punto di partenza per l'analisi del proverbio è l'assunzione che il proverbio sia una frase. A questo punto, ci dobbiamo chiedere di che tipo di frase si tratti. Se necessario, possiamo confrontare il proverbio con altre grandezze commensurabili, ossia con altre frasi.

1.4.1. Frase ON-sentencieuse

In primis, il proverbio deve essere una cosiddetta frase *On-sentencieuse*. Ne parla Anscombe (2000: 11-12) a proposito del raffronto tra i proverbi e gli altri tipi di fenomeni sentenziosi. In altre parole, l'autore si domanda come si possa distinguere tra il proverbio e, ad esempio, l'aforisma o l'apoftegma. In realtà, l'elenco dei fenomeni sentenziosi potrebbe continuare ancora per molto. Pensiamo alle paremie individuate all'interno del *Refranero Multilingüe*¹⁹, una grande raccolta elettronica di proverbi e frasi sentenziose spagnole, che non solo dà conto dell'esistenza di queste paremie in lingua spagnola, ma ne fornisce adeguate definizioni, classificazioni tipologiche e traduzioni in altre lingue. Si tratta di un progetto di ampio respiro portato avanti dal Gruppo di Ricerca *Frasesología y Peremiología*, diretto da Julia Sevilla Muñoz e M. I. Teresa Zurdo Ruiz-Ayúcar.

Si vedano, qui di seguito, i diversi tipi di paremie individuate in questa raccolta di proverbi: aforisma, apoftegma, stereotipo, dialogismo, frase fatta, frase proverbiale, locuzione proverbiale, massima, paremia, proverbio colto,²⁰ proverbio popolare,²¹ proverbio dialogato, proverbio meteorologico, proverbio temporale, sentenza. Non è facile districarsi in questa intricata selva di generi di *forme brevi* (per un'analisi più completa delle cosiddette forme brevi, si rimanda a Schapira, 1999: 48-53), che vengono frequentemente confusi (cfr. Montreynaud et al., 1989: X), anche se il loro confronto potrebbe essere di aiuto nella comprensione di che cosa siano i proverbi.

La prima osservazione che potrebbe emergere dall'analisi di questa terminologia sentenziosa, riprendendo il sintagma "*terminologie sentencieuse*" utilizzato da Anscombe (2000: 8), è che nello stesso *Refranero Multilingüe* non sono chiari i confini tra un genere e l'altro o tra un'etichetta e l'altra. Questa tendenza non è presente solo in questo studio, ma rispecchia una confusione terminologica ben più ampia e generalizzata, probabilmente perché non esiste ancora oggi un accordo sui diversi fenomeni gnomici, proprio come sottolinea lo stesso Anscombe (2000: 8). L'autore, ad esempio, ricorda come fenomeni sentenziosi analoghi siano posti sotto le più svariate categorie a seconda degli autori che se ne occupano. Anche María Jesus Barsanti Vigo (2006: 198) ricalca una certa ambiguità in campo paremiologico. Secondo la studiosa, etichette come detti, modi di dire, proverbi, locuzioni proverbiali, frasi fatte, sentenze, aforismi, stereotipi, adagi, apoftegmi, massime sono presenti

¹⁹ Per maggiori informazioni, si rimanda al sito internet: <http://cvc.cervantes.es/lengua/refranero/Default.aspx>.

²⁰ Il termine scelto in spagnolo è *proverbio*. Per una disamina della differenza tra *proverbio* e *refrán* cfr. Sevilla Muñoz (2000: 101).

²¹ Il termine scelto in spagnolo è *refrán*. Per una disamina della differenza tra *proverbio* e *refrán* cfr. Sevilla Muñoz (2000: 101).

nei dizionari e nella comunicazione quotidiana senza che gli studiosi ne abbiano previamente fissato il significato preciso. Dal canto nostro notiamo, invece, che tali etichette sono molto approssimative e, in diversi casi, errate. Basti pensare che frequentemente ciò che si definisce *frase fatta* in realtà non è una *frase* ma solo un *costituente*.

Per spiegare la differenza tra i proverbi e gli altri fenomeni gnomici, e al contempo reperire la prima delle condizioni necessarie perché una frase sia un proverbio, dobbiamo distinguere tra frasi *L-sentencieuse* e frasi *ON-sentencieuse*. Per frasi *L-sentencieuse* si intendono quelle frasi sentenziose il cui enunciatore primo è conosciuto. Questo significa che sono frasi che vengono ripetute da un numero N di locutori in N situazioni ma che posseggono un autore, che corrisponde al locutore che per primo pronunciò quella data frase. A titolo di esempio, può essere frase *L-sentencieuse* qualunque tipo di massima, aforisma, sentenza, *inter alia*, il cui autore, o enunciatore primo, è noto alla comunità di parlanti. Al contrario, si considerano frasi *ON-sentencieuse* quelle che non hanno un enunciatore primo (vale a dire un autore) riconoscibile o noto ma che appartengono alla saggezza popolare, all'osservazione quotidiana e a un sapere condiviso. Esempi di frasi *ON-sentencieuses* sono sia frasi idiomatiche come *Un ange passe*, definite da Anscombe frase *ON-sentencieuse épisodique*, sia *Une hirondelle ne fait pas le printemps*, frase *ON-sentencieuse typifiante à priori*. L'enunciatore primo di queste frasi non è definito o non è più noto alla comunità²²: nella terminologia di Anscombe un *ON-énonciateur*. Pertanto, sono incompatibili, da un lato, con specificazioni del tipo *Come ha detto X* (ma anche *Come disse una volta un tale*) ma perfettamente combinabili con commenti del tipo *Come si dice*, *Come si suol dire*, ecc. Certo, è molto probabile che anche queste frasi abbiano avuto un autore primo, ma è come se costui (o costoro) avesse ceduto alla comunità linguistica il diritto di applicare il principio generale enunciato dalla frase ai casi particolari. L'autore diventa così indefinito e non più riconoscibile.

Ciò non toglie che nella comunicazione spontanea non si possa indicare un individuo ben preciso come autore del proverbio. Questo è quanto accade nel seguente esempio: *Mia*

²² Da sottolineare il valore fondamentale del giudizio dell'intera comunità: il fatto che, ad esempio, per uno o pochi individui non sia più noto l'autore di un aforisma, non ne fa, di conseguenza, una frase *ON-sentencieuse*. Se, invece, è tutta la comunità a non riconoscere più dietro a un antico aforisma l'autore o enunciatore primo dello stesso, allora è possibile considerare l'aforisma una frase *ON-sentencieuse*. Allo stesso modo, se secondo un parlante un proverbio è un'invenzione di una persona a lui nota, e quindi assimilabile a una frase *L-sentencieuse*, non per questo il proverbio cessa di essere considerato tale dalla comunità di parlanti. Un esempio di questo si trova nel proseguio di questo § 1.4.1.

*nonna mi diceva: meglio tardi che mai.*²³ A prima vista, sembrerebbe che la frase *Meglio tardi che mai* sia una citazione con un autore primo ben riconoscibile, ossia una frase *L-sentencieuse*, ed è solamente la nostra esperienza in campo paremiologico, nonché il giudizio dell'intera comunità linguistica, a farci capire che in realtà si tratta di un proverbio.

Con questa prima classificazione, siamo in grado di separare le forme sentenziose aventi un autore primo riconoscibile dai proverbi. Esempi di frasi *L-sentencieuses* sono, tra gli altri, gli aforismi, gli apoftegmi, le massime, le sentenze. Il prototipo, invece, delle frasi *ON-sentencieuses*, è senz'altro il proverbio, il quale si rivela come una frase il cui autore primo è scomparso dalla scena e "ha ceduto i suoi diritti" all'intera comunità.

1.4.2. Frasi generiche

Continuando a restringere il campo delle frasi candidate a diventare proverbi, troviamo il tema della genericità, presente in maniera trasversale in molti studi legati alla paremiologia. Tale nozione è presente in alcuni lavori della Scuola Geoparemiologica Italiana. In seno ad essa, si sostiene che il *valore paremiologico* sia un tratto necessario alla funzione proverbiale. Si definisce valore paremiologico "l'ampio spettro semantico da cui si trae analogicamente il significato peculiare" a un determinato contesto situazionale (Cervini et al., 1999: 181). Il valore paremiologico, sempre secondo la Scuola Geoparemiologica Italiana, non è solamente tipico dei proverbi non compositivi (definiti *impliciti*), ma anche di quelli compositivi (definiti *espliciti*), in quanto anche i proverbi compositivi godono di una certa generalizzazione semantica in ragione della loro genericità referenziale (Cervini et al., 1999: 181).

Una delle prove che potremmo applicare ai proverbi per testare la loro genericità è l'attualizzazione. In altre parole, se attualizzando un proverbio questo cessa di essere tale, ciò significa che la genericità è un tratto essenziale del proverbio. Proviamo a modificare un proverbio, come potrebbe essere *Chi si loda s'imbroda*, e ad attualizzarlo dicendo *Luca si loda e s'imbroda*, oppure *Di solito quando Luca si loda s'imbroda*. Produciamo una frase che potrebbe ricordare un proverbio ma che non è evidentemente un proverbio (si veda altresì Perrin, 2000: 70). Se la sola perdita della proprietà di genericità impedisce a un proverbio di continuare a essere considerato tale, significa che ci troviamo probabilmente di fronte a una caratteristica propria e necessaria dei proverbi.

²³ Esempio estrapolato da una conversazione spontanea avvenuta a ottobre del 2015, in provincia di Ancona.

Giunti alla consapevolezza che i proverbi devono essere frasi generiche, potremmo domandarci che tipo di frasi generiche debbano essere e a quale realtà faccia riferimento tale genericità. In altri termini, che cosa sono in grado di denominare i proverbi?

1.4.3. Proverbi e denominazione

Kleiber ha dedicato molti dei suoi studi ai proverbi intesi come frasi generiche e considerati come denominazioni di altre cose. Ha anche suscitato perplessità e critiche, le quali gli hanno permesso di specificare meglio che cosa intenda per denominazione dei proverbi. Pensiamo allo scambio di opinioni, a distanza ma assai vivace, tra Geroges Kleiber²⁴ e Christine Michaux²⁵ proprio su questo argomento.

In effetti, come ammette lo stesso Kleiber (2000), sarebbe errato considerare questo paradigma sulla falsariga della relazione tra un sostantivo comune e il concetto che esso denota. I proverbi non fanno riferimento a semplici concetti, bensì a situazioni. La nozione di denominazione non è da prendere nel senso di *noun*, ma nel senso di *name*. Secondo Kleiber, i proverbi sono unità polilessicali codificate con una certa rigidità a livello formale e un'elevata stabilità semantica (Kleiber, 2000: 40). Tale fissità rende il senso dei proverbi precostituito e quindi convenzionalmente valido per qualsiasi parlante, diventando parte del codice linguistico comune (Kleiber, 2000: 40 e Norrick, 1985: 1).

Cette vertu de *name* lui (au proverbe) permet de catégoriser, c'est-à-dire de ranger ou rassembler dans la catégorie dont il est la dénomination, des occurrences particulières qui les vérifient. Il n'y a donc pas de rapprochement ou d'assimilation à effectuer avec les noms communs, même si ceux-ci peuvent passer à plus ou moins juste titre pour les prototypes des dénominations.

(Kleiber, 2000: 40)

Sarebbe un errore trascurare la natura polilessicale, e soprattutto frastica, dei proverbi. Non a caso, essi sono in grado di denominare altre cose mantenendo intatto il loro carattere di frase. Come suggerisce Kleiber (2000: 41), se i sostantivi, i verbi, gli aggettivi sono denominazioni che mantengono inalterati i propri attributi specifici delle categorie grammaticali a cui appartengono, anche i proverbi sono denominazioni che rispettano (o forse rispecchiano) la loro realtà di frase, e più precisamente di frase generica. Questi indici

²⁴ Cfr. Kleiber (2000: 40).

²⁵ Cfr. Michaux (1999).

consentono di cogliere il referente denominato dai proverbi, che, in questo caso, è una situazione o uno stato di cose generico. Non sono, dunque, denominazioni di oggetti o di entità per così dire nominali, bensì di stati di cose generali, abituali, consuetudinari. Non si occupano della contingenza, dell'accidentale, delle occorrenze specifiche che riguardano gli individui o gli eventi, bensì del campo gnomico o *law-like* delle frasi generiche, in cui le relazioni espresse sono, per così dire, diventate indipendenti dalle singole situazioni concrete, pur mantenendo con queste un legame diretto. Come sostiene Kleiber (2000: 41), "*les proverbes en tant que phrases génériques expriment ainsi des régularités structurantes et non des assertions sur des faits particuliers*".

Proseguendo nella trattazione della denominazione dei proverbi, possiamo domandarci di che tipo sia il senso dei proverbi. In altri termini, che natura ha generalmente il senso della macrocategoria dei proverbi? Riprendendo un punto proposto da Kleiber (2000: 42), da un lato i proverbi sono unità polilessicali codificate e, dall'altro, sono frasi generiche. Che tipo di senso genera tale binomio?

Non tutti i ricercatori sono concordi riguardo alla sensatezza di cercare un senso generale da applicare a tutta la vasta categoria dei proverbi. Per molti, infatti, i proverbi sono talmente diversi e distanti tra di loro, talmente eterogenei dal punto di vista semantico, che è impossibile indicare a priori quale sia la tipologia di senso che bisogna attribuire loro. Probabilmente questo scetticismo ha una certa fondatezza, specialmente considerando che i proverbi, essendo frasi, possono diventare enunciati e veicolare messaggi contingenti determinati dal contesto comunicativo. Tuttavia, ha senso chiedersi se non ci sia qualche elemento semantico trasversale a tutti i proverbi che li renda simili anche dal punto di vista di come e cosa possono denominare.

Kleiber (2000: 44) propone, inoltre, di riflettere sul fatto che, seppure non nella totalità dei casi, ma almeno con una certa frequenza, sia possibile comprendere correttamente il significato idiomatico o paremiologico di un proverbio sconosciuto. L'obiezione lecita e giusta che si può muovere a questa considerazione è che molto spesso non si riesce a comprendere correttamente un proverbio mai sentito in precedenza (considerando anche che ciò potrebbe avvenire anche nel caso di proverbi noti). Tuttavia, come sottolinea Kleiber, in linea con quanto emerge dagli studi di altri autori quali Gibbs (1994: 309-317) e Schapira (1999: 58-60), anche in questo ultimo caso si nota che i parlanti non interpretano a caso il senso dei proverbi sconosciuti, ma usano schemi fissi. Continuando il nostro "gioco" delle

differenze, vediamo quali sono le principali discrepanze tra i proverbi e le frasi generiche. Per fare questo, analizzeremo, in ordine, i tre piani su cui si sviluppa la semantica dei proverbi: il significato compositivo, il significato idiomatico, che nel caso dei proverbi preferiamo definire paremiologico per non confonderlo con quello delle espressioni idiomatiche, e il messaggio veicolato.

Partendo dal significato compositivo, notiamo che proverbi e frasi generiche condividono la capacità di poter assumere qualsiasi tipo di significato compositivo. Entrambe le categorie di frasi, infatti, possono riferirsi composizionalmente tanto a realtà che riguardano l'uomo quanto a qualsiasi realtà o fatto che non abbia nulla a che vedere con l'uomo. Pensiamo alle frasi generiche *L'uomo è un essere mortale* oppure *L'oro è giallo* e ai proverbi *Tale padre, tale figlio* oppure *Non c'è rosa senza spine*. Notiamo che il significato compositivo di queste quattro frasi non è prevedibile e spazia in ambiti molto diversi.

Passando al secondo grado di analisi, relativo al significato idiomatico e paremiologico, osserviamo il diverso comportamento delle frasi generiche e dei proverbi. Se prendiamo nuovamente la frase generica *L'oro è giallo*, ma anche *Le scimmie mangiano le banane*, notiamo che non è possibile alcuna lettura in senso umano. Al contrario, al proverbio *Non c'è rosa senza spine* è certamente possibile e quasi scontato fornire una lettura in senso umano. Questo è un punto cruciale, secondo Kleiber (2000: 45): un tipo di frase generica come quella riportata non può accedere allo status di proverbio in quanto non può essere interpretata in senso umano. Anche se fosse trasposta in senso metaforico, non riguarderebbe comunque l'uomo, il che invece accade per il proverbio.

Arriviamo ora al terzo stadio della nostra analisi relativo al messaggio che può veicolare un proverbio. Sappiamo bene che è impossibile stabilire a priori il messaggio che trasmette una qualsiasi frase, dato che non si possono identificare a priori le variabili contestuali che necessariamente influiscono sulla definizione del messaggio di una frase. Tuttavia, il piano del messaggio è funzionale alla nostra analisi. Si potrebbe, infatti, sostenere che il significato idiomatico possa essere influenzato o aiutato dal contesto. È bene da subito chiarire questo punto controverso: anzitutto, non si può confondere il piano semantico (significato paremiologico) con il piano pragmatico (messaggio veicolato). Il significato paremiologico, infatti, prescinde dal messaggio veicolato, in quanto viene prima della formazione del messaggio. Tale significato, infatti, viene attribuito unanimemente al proverbio ancor prima che questo si sia fatto enunciato. I parlanti sanno che cosa significa un

proverbio, così come i dizionari ne forniscono il corretto significato paremiologico. Ecco che il significato paremiologico non è influenzato dal messaggio contingente del proverbio. Pertanto, quando si parla di restrizione semantica all'uomo si intende esclusivamente il piano del significato paremiologico. Non a caso, anche Kleiber (2000: 45) sottolinea che i proverbi sono interpretabili in chiave umana anche in assenza di contesto.

Queste considerazioni nascono da una semplice osservazione, ma di grande impatto e rilevanza: a proposito dei proverbi, vi è una vera e propria competenza, in virtù della quale siamo in grado di distinguere ciò che è semanticamente un proverbio da ciò che non lo è. Come sottolinea Kleiber (2000), se i parlanti hanno una competenza legata all'uso e al riconoscimento dei proverbi, questa è per forza legata alla condizione semantica che ogni frase generica deve soddisfare per poter diventare o essere considerata un proverbio: la restrizione semantica all'uomo. Pertanto, una frase generica può diventare un proverbio o, quanto meno, essere candidata ad assumere lo status di proverbio se possiede questo tratto semantico "umano" nel suo significato paremiologico.

A questo punto è opportuno aprire una parentesi, non certo di minore importanza, sui quei proverbi che riguardano, tra gli altri, il tempo atmosferico, le attività rurali, come le semine e i raccolti, l'andamento ciclico delle stagioni e le loro caratteristiche principali. Tali proverbi, che costituiscono un'alta percentuale della totalità dei proverbi, assumono un tono alquanto didattico o "scientifico" (per quanto questo sia possibile, trattandosi di saggezza popolare) al fine di spiegare alcuni fenomeni meteorologici molto concreti o dare consigli su quando e come effettuare determinati lavori e attività. Sembrerebbe, dunque, che questa categoria di proverbi non rispetti la condizione generale dei proverbi stessi, ossia quella di avere una semantica ristretta agli uomini. In realtà, le cose non stanno proprio così, almeno secondo Kleiber (2000: 46) che, riportando il pensiero di altri studiosi come Anscombe e Schapira, sottolinea come qualsiasi questione meteorologica riguardi l'uomo. Per inquadrare meglio tale affermazione, dobbiamo pensare alla società che ha dato alla luce tali proverbi, una società sicuramente rurale, il cui sostentamento si basava sui prodotti della terra e della natura. Ecco perché il clima è in grado di per sé, in una società come quella delineata, di influire profondamente sulla condotta, sui pensieri e sulle azioni dell'uomo. In realtà, anche oggi, seppure con tutti i cambiamenti che differenziano la nostra società occidentale contemporanea da quelle rurali del passato, il clima ha un enorme peso sulle nostre vite e sui

nostri comportamenti. Lo stesso vale per quei proverbi prescrittivi che regolavano le attività delle società preindustriali.

Tuttavia, questa affermazione non ci convince appieno, poiché sembra fornire una giustificazione di quanto il significato compositivo dei proverbi possa comunque riferirsi all'uomo e non di come da un significato compositivo apparentemente inconciliabile con l'uomo si riesca a passare a un significato paremiologico che prenda in considerazione la natura umana.

Ciò che occorre tenere maggiormente in considerazione riguardo alla semantica dei proverbi è il fatto che sia i proverbi in generale sia i proverbi relativi al meteo garantiscono un principio generale di ragionamento che fa passare da un livello (nozione meteorologica) a un altro (relativo all'uomo), denotando un *topos* che permette tale passaggio. Kleiber (2000: 46) arriva alla conclusione che i proverbi meteorologici riguardano direttamente i fenomeni naturali e solo indirettamente l'uomo, mentre gli altri proverbi sono frasi generiche che riguardano direttamente l'uomo (Kleiber, 2000: 46). Questa formulazione ancora non ci convince, dato che ci sembra nuovamente mescolare i diversi piani su cui abbiamo sviluppato l'analisi semantica dei proverbi.

Pertanto, indipendentemente dal contenuto semantico del significato compositivo, i referenti dei proverbi possono essere qualsiasi e possono mostrare un'attinenza maggiore o inferiore (o persino nulla) con la sfera umana.

Dal punto di vista del contenuto paremiologico, i proverbi, per essere considerati tali, devono avere la capacità di riferirsi all'uomo (intendendo, per uomo, la sua natura, i suoi comportamenti e quanto sia pertinente con la sua vita). Ciò non toglie che non si possa utilizzare un proverbio in virtù del suo significato compositivo e, quindi, non ristretto semanticamente all'uomo. Ciononostante, in questo caso saremmo di fronte a un fenomeno di *déproverbialisation*, in virtù del quale il proverbio abbandonerebbe, occasionalmente e reversibilmente, il suo status proverbiale per assumere lo status di frase libera e seguire i normali principi della semantica come, ad esempio, la compositività. Prendiamo i seguenti proverbi: *Il bel tempo non dura sempre; Dopo il lampo viene il tuono; Dopo la neve, buon tempo ne viene; Fango di maggio, spighe d'agosto*. Sebbene essi sembrano normali frasi generiche, in realtà possono essere considerati veri e propri proverbi, non solo in quanto è possibile una loro trasposizione metaforica, ma anche perché questa è ristretta all'uomo. Infatti, se ci si concentra sul significato paremiologico di questi proverbi, notiamo che esso

non è volto a descrivere soltanto alcuni fenomeni atmosferici, bensì riguarda la vita delle persone oppure circostanze in cui l'uomo si può trovare. Prendiamo come esempio il proverbio *Il bel tempo non dura sempre*, che viene definito dal *Grande dizionario dei proverbi italiani* (Guazzotti e Oddera, 2006) come segue: “può essere inteso in senso letterale o figurato: in quest'ultimo caso vale come ammonimento a non dare per scontato il favore della sorte”. Nel dizionario di Lapucci (2006) troviamo una variante del proverbio in esame, ossia *Buon tempo e mal tempo non durano tutto il tempo*, che viene spiegata nella seguente maniera: “La bella e la brutta stagione non sono stabili, si alternano come i periodi fortunati o infausti della vita”. Si noterà che nella prima definizione offerta, tratta da Guazzotti e Oddera (2006), si fa esplicito riferimento sia al significato compositivo sia a quello paremiologico del proverbio, quasi come a giustificare il fatto che questo proverbio meteorologico funzioni semanticamente come un qualunque proverbio. Anche in Lapucci (2006) si menziona il significato compositivo del proverbio, ma in maniera meno prorompente rispetto al primo dizionario e soltanto come analogia per spiegare il significato paremiologico del proverbio. Dal nostro punto di vista, non è necessario sottolineare che il proverbio può essere impiegato sia compositivamente sia in funzione del significato paremiologico: questo sta nella natura del proverbio stesso e non riguarda il livello della semantica del proverbio, bensì il livello successivo dell'uso del proverbio, ovvero il livello pragmatico. Qualsiasi proverbio, anche il comunissimo *A caval donato non si guarda in bocca*, che non viene praticamente mai impiegato secondo il senso compositivo, può essere utilizzato in funzione del significato compositivo, sebbene questa informazione manchi dal dizionario di Guazzotti e Oddera (2006). Quanto ci può suggerire questo zelo di specificare la modalità di uso del proverbio dipende probabilmente dall'alta frequenza d'uso dei proverbi meteorologici secondo il loro significato meteorologico e non in base a quello paremiologico²⁶. Crediamo che si tratti di una

²⁶ Proprio per questa ragione, secondo molti paremiologi i proverbi meteorologici non possono essere considerati proverbi, dato che non si riferirebbero nella maggioranza dei casi all'uomo. In questo studio, non ci fermeremo molto su questo tema, che pure è di grande peso nell'attuale ricerca in ambito paremiologico. Per quanto riguarda lo studio in oggetto, e specialmente i risultati estratti da *corpora* linguistici che verranno analizzati nella seconda parte di questo lavoro, notiamo che è vero che i proverbi meteorologici vengono impiegati con altissima frequenza solo in virtù del loro significato compositivo. Proverbi come *Rosso di sera, bel tempo si spera; Per la Candelora, dell'inverno siamo fora; Cielo a pecorelle, pioggia a catinelle*; ecc. vengono effettivamente utilizzati quasi sempre secondo il loro significato compositivo e non lasciano molto spazio a un'interpretazione in chiave umana. Tuttavia, è innegabile il loro attaccamento alla natura umana (la speranza è un'attitudine prettamente umana, la Candelora è una festa religiosa e quindi istituita dall'uomo, l'individuazione di tante piccole pecore in cielo è una fantasia puramente umana). Inoltre, siamo convinti che, anche solo per giochi ironici, si possono trovare casi in cui si impiegano in virtù del loro significato paremiologico. Pensiamo al gioco di parole *Rosso di sera, mal di testa la mattina* (esempio tratto dal corpus itWaC; per maggiori

tendenza secondo la quale si usano questi proverbi secondo il loro senso compositivo ma che non ne annulla a priori il significato paremiologico.

Anche numerosi altri autori, quali Schapira (1999: 69-73) e Lakoff e Turner (1989: 174-175), solo per citarne alcuni, mettono in luce che tutti i proverbi, compresi quelli meteorologici, riguardano l'uomo. In particolare, per Lakoff e Turner (1989) la capacità di indicare l'essere umano è una caratteristica essenziale del proverbio. Ad esempio, il proverbio *Big thunder, little rain* può riferirsi esclusivamente ad un fenomeno atmosferico, ovvero, utilizzando gli estremi della nostra analisi, attivare solamente il significato compositivo. Ciononostante, in questo caso i parlanti capirebbero che la frase non verrebbe più usata in qualità di proverbio. Lo stesso proverbio, come mostrato poco sopra per gli altri proverbi metaforici italiani, potrebbe invece mettere in campo il proprio significato paremiologico e assumere lo stesso significato paremiologico dei proverbi italiani: *Tanto rumore per nulla*; *Tanto fumo e poco arrosto*; *Partoriscono le montagne e nasce un topolino*; *La montagna ha partorito un topolino*²⁷.

In definitiva, una *conditio sine qua non* perché una frase generica sia o possa diventare un proverbio è non solo avere un senso paremiologico nei termini spiegati in questo paragrafo, ma anche poter riferirsi in virtù di questo senso, e indipendentemente dal significato compositivo, all'uomo.

1.4.4. La parabola del proverbio

Appurato che è pertinente parlare di significato paremiologico relativamente ai proverbi, bisogna ora chiedersi perché, sebbene con diversi gradi di frequenza a seconda dei tipi di proverbi (se di tipi di proverbi si può parlare), è possibile assegnare di volta in volta ai proverbi un'interpretazione letterale o un'interpretazione non letterale.

Entrano in campo, a questo punto, diverse variabili: anzitutto, il rapporto tra il significato compositivo e il significato paremiologico e, successivamente, il rapporto tra il significato paremiologico e il messaggio del proverbio enunciato in situazioni comunicative.

Il rapporto tra il significato compositivo e quello paremiologico di un proverbio è governato principalmente da ragioni metaforiche e metonimiche. Ciò che presenta compositivamente un proverbio è una sorta di allegoria che permette il passaggio dal

informazioni sul corpus, cfr. § 2.3) sebbene questo proverbio sia modificato, la modifica non avviene a caso ma con precisi canoni che riguardano precisamente l'uomo e, in questo caso, i suoi vizi. Pertanto, siamo portati a considerare anche i proverbi meteorologici veri proverbi.

²⁷ Proverbi tratti da Lapucci (2006).

significato compositivo a quello paremiologico. Prendiamo come esempi i proverbi *A caval donato non si guarda in bocca* e *Il bel tempo non dura sempre*. Essi sono rispettivamente metafore di regali che non devono essere criticati e di eventi fortunati dalla durata limitata. Vi sono modalità ricorrenti che consentono di passare dal significato compositivo a quello paremiologico. Come sottolineato nel precedente § 1.4.3, i proverbi denotano situazioni generiche che hanno attinenza con l'uomo. Si è già analizzato in dettaglio il valore del riferimento alla sfera umana nella quasi totalità dei proverbi. Concentriamoci ora sul valore della genericità. Se analizziamo gli esempi appena riportati, notiamo che il senso della direzione tra significato compositivo e significato paremiologico è unico: si va dalla situazione particolare (cavallo/non guardare i denti; bel tempo/durata limitata) a una situazione generale (regalo/non criticare; evento positivo/durata limitata). Il significato paremiologico opera una sorta di astrazione del significato compositivo.

Dalla fase di astrazione del significato paremiologico si passa successivamente alla fase di utilizzo nella comunicazione del proverbio e quindi allo stadio di creazione del messaggio. Il proverbio è una frase (che è la soglia minima del proverbio, sebbene, come già più volte sottolineato, nulla impedisca al proverbio stesso di essere un testo). Pertanto, nel momento in cui, diventato enunciato, deve essere interpretato, seguirà alcuni dei principi fondamentali dell'interpretazione delle espressioni linguistiche. Si è già detto che non ha senso parlare di proverbi letterali o proverbi non letterali, in quanto l'essere letterale o non letterale non è una proprietà intrinseca del proverbio, bensì una caratteristica legata alla sua interpretazione. Lo stesso discorso vale per le espressioni linguistiche in generale: anche in questo caso, non esistono espressioni linguistiche letterali o non letterali, esistono interpretazioni letterali o non letterali di esse. La scelta dell'interpretazione di un'espressione è motivata da una co-occorrenza contingente di fattori contestuali che rispondono a un mero criterio di coerenza (cfr. Prandi, 2004: 13). Per la medesima ragione, l'interpretazione non letterale è motivata da una catena di inferenze che si sviluppa nel caso in cui il significato di un'espressione e un messaggio non coincidano, il che si verifica quando l'interpretazione letterale manca di coerenza con i dati contestuali e cotestuali. Sebbene l'interpretazione letterale sia da considerarsi quella preferita nella comunicazione da parte del destinatario del messaggio, quando essa non è rilevante, per via di un divario tra un significato linguistico e un messaggio, il destinatario stesso deve ricorrere a una catena di inferenze per mantenere un atteggiamento cooperativo e comprendere il messaggio dell'enunciato. Capiamo, dunque,

perché la relazione tra significato e messaggio è sempre di natura contingente e legata a fattori contestuali. Questo è spiegato dal fatto che la relazione tra significato e messaggio è di tipo indicale. L'indice, al contrario del significato, non è governato da un codice, bensì da un insieme di fattori occasionali e co-occorrenti che sono imprevedibili. Ecco perché il messaggio deve essere interpretato sulla base di informazioni co-occorrenti condivise (cfr. Prandi, 2004: 19).

Il carattere contingente e non predittivo del messaggio non deve far dimenticare che per interpretare un messaggio bisogna identificare il valore di un indice all'interno di un campo, ossia stabilire una relazione tra un segnale strutturalmente stabile e un messaggio contingente. Ecco che ha senso chiedersi quali siano i parametri con cui si è soliti interpretare un proverbio enunciato in una situazione comunicativa.

Effettivamente, il proverbio dopo aver subito una fase di astrazione a livello del suo significato paremiologico, perde quella genericità che aveva acquisito in questo stadio per assumere nuovamente un carattere di specificità. Esso, infatti, denota una situazione particolare direttamente determinabile dal contesto della situazione comunicativa in cui viene enunciato, mantenendo intatta la relazione con la situazione generica di cui è denominazione a livello di significato paremiologico, ma applicando tale struttura alle situazioni specifiche che emergono dai contesti contingenti.

Considerando quanto fin qui sostenuto in un'altra prospettiva, il proverbio segue l'andamento di una parabola: parte da una situazione particolare, quella direttamente indicata a livello del suo significato compositivo, per proseguire il percorso ascendente fino al punto culminante, rappresentato dal significato paremiologico del proverbio che denomina la situazione generica del proverbio, per poi proseguire il percorso scendendo verso il punto finale, rappresentato dal singolo caso attualizzato e specifico a cui si applica il proverbio.

1.5. Proverbi, *déproverbialisation* e *détournement*

Come si è già più volte accennato, i proverbi, oltre ad avere diverse varianti attestate, possono frequentemente essere oggetto di modifiche, manipolazioni, variazioni e stravolgimenti compiuti dal parlante a proprio piacere. Tali cambiamenti interessano diversi livelli: quello lessicale, ma anche quello semantico e stilistico. Charlotte Schapira, nel suo articolo intitolato "*Proverbe, proverbialisation et déproverbialisation*" (Schapira, 2000: 81-97), illustra questi fenomeni, che elenca, assieme ad altri, nella sezione intitolata "*La déproverbialisation*". Molto interessante, prima di affrontare lo studio delle manipolazioni

operate sui proverbi, risulta il processo che l'autrice, in linea con altri ricercatori, denomina *déproverbialisation*. Se la *proverbialisation* indica il divenire proverbio da parte di una certa espressione o comunque l'atto con cui una frase si fissa nell'uso in qualità di proverbio, la *déproverbialisation* non indica il processo contrario con cui un proverbio cessa di essere considerato e utilizzato come tale (Schapira, 2000: 93), bensì un procedimento con cui, in un'occorrenza puntuale, lo stesso proverbio viene restituito al discorso libero, ritornando a essere una proposizione semplice, come se fosse stata creata liberamente nel discorso (Schapira, 2000: 93).

Entriamo, ora, nel merito di tali manipolazioni e variazioni. Il primo esempio di *déproverbialisation* proposto da Schapira (2000: 93) è la negazione del messaggio proverbiale. Una delle principali caratteristiche attribuite solitamente ai proverbi è il fatto che essi siano il prodotto della saggezza delle genti, cosa che permette loro di avere un'elevata autorevolezza e di non essere contraddetti. La stessa Schapira (2000: 86) sottolinea che il messaggio proverbiale, una volta fissatosi nell'uso di una comunità linguistica, viene considerato dai parlanti della stessa comunità un dato di fatto, un pensiero stereotipato che entra a pieno titolo nella *doxa*, quel bagaglio di opinioni comuni condivise dai membri di una stessa comunità linguistica in un momento dato della sua storia. La studiosa ricorda la somiglianza dei proverbi con i luoghi comuni e le leggi scientifiche, che condividono l'essere opinioni acquisite e dati di fatto. Tuttavia, almeno dal punto di vista linguistico, si nota che i proverbi possono differire dalle leggi scientifiche, trattandosi, infatti, di leggi umane. Ecco perché se da un lato le leggi scientifiche sono incompatibili con i verbi di opinione e non sono atualizzabili, come è possibile osservare negli esempi seguenti:

- (6) a. * Trovo che due rette parallele non si incontrino mai.
b. * L'anno scorso / l'anno prossimo due rette parallele non si incontravano / non si incontreranno mai.

dall'altro lato i proverbi possono essere compatibili in certi casi e in certi contesti con verbi di opinione e posso essere atualizzati:

- (7) a. Trovo che non ci sia rosa senza spine.
b. Mio nonno racconta che ai suoi tempi il mattino aveva l'oro in bocca.

Il fatto che talvolta sia possibile atualizzare i proverbi non significa che questi non appartengano al bagaglio di opinioni condivise di una comunità linguistica, quanto piuttosto

che è possibile che essi perdano la loro caratteristica di verità atemporale, innescando così un processo di *déproverbialisation*.

Schapira (2000: 93) prosegue la sua trattazione sulla *déproverbialisation* mostrando come il fatto di portare dei contro esempi al proverbio sia un ulteriore strumento di *déproverbialisation*. Tuttavia, la negazione dei proverbi non è semplice come potrebbe sembrare. Anzitutto, perché come sottolinea Schapira (2000: 93), gli esempi presenti in corpora linguistici di proverbi negati sono molto scarsi e, per studiare tale fenomeno, si usano quasi sempre frasi costruite ad hoc dai linguisti. Trascurando il fatto che i dati su cui si è lavorato fino ad oggi siano artificiali e non autentici, cerchiamo di analizzare due esempi proposti da Schapira (*ibid.*).

- (8) a. Comme dit le proverbe, qui aime bien châtie bien. – ? Non, car il y a des gens qui passent tout à ceux qu'ils aiment.
- b. Comme dit le proverbe, qui aime bien châtie bien. – (Oui / Bien sûr / Certes) mais je connais des gens qui passent tout à ceux qu'ils aiment.

Nell'esempio (8a), seguendo alla lettera quanto indicato da Schapira nel proprio lavoro, è stato inserito un punto interrogativo per segnalare una certa esitazione quanto all'accettabilità dell'esempio. In effetti, almeno per ora, non vi sono ricorrenze in contesti reali di negazioni così nette, segnalate da un 'no' secco (in seguito, analizzando approfonditamente il corpus itWaC, si vedrà se emergeranno casi di questo tipo). Anche l'esempio (8b) è stato forgiato *ad hoc*, ma sembra più accettabile. In effetti, non si confuta immediatamente il proverbio, bensì si conferma *in primis* che l'enunciato prodotto ha lo status di proverbio (e, quindi, che è foriero di una verità generale) ma si segnala il disaccordo riportando un contro esempio che proviene dall'esperienza particolare dell'interlocutore. In altri termini, non si confuta la validità del proverbio *in toto*, ma si portano esempi *ad hoc* che mostrano una devianza dalla norma. Si potrebbe affermare che il proverbio viene accompagnato da un'eccezione che, in realtà, conferma la regola. Questo tema è legato al discorso delle frasi generiche: il fatto che si portino contro esempi per dimostrare che la realtà può discostarsi dalla regolarità non falsifica la frase generica. Se si afferma: "Le scimmie mangiano le banane, ma non Cheetah" (Anscombe, 1994: 103), non si mette in discussione la validità della frase generica *Le scimmie mangiano le banane*. Parimenti, se si afferma: "Questa rosa non ha spine", non si invalida il proverbio *Non c'è rosa senza spine*. Tornando al nostro esempio (8b), si è detto che la risposta è più accettabile di (8a) in quanto con essa si

riconosce dapprima lo status di verità generale del proverbio, portando successivamente contro esempi per dimostrare la sua non applicabilità alla totalità dei casi. Come evidenzia Schapira, siamo in presenza di due fenomeni concomitanti e simultanei di *proverbialisation* (segnalata da *Oui / Bien sûr / Certes*) e *déproverbialisation* (*Mais je connais des gens qui passent tout à ceux qu'ils aiment*).

Questo potrebbe spiegare perché un proverbio sia difficilmente introdotto da un verbo di opinione alla forma affermativa, mentre lo stesso verbo alla forma negativa in posizione di introduzione al proverbio sia maggiormente accettabile. Se, infatti, è poco coerente utilizzare tali verbi alla forma affermativa per introdurre un proverbio (inteso come stereotipo) e mostrarlo come un pensiero originale, non è da escludere un verbo di opinione alla forma negativa che mostra semplicemente il dissenso nei confronti dell'idea ricevuta.

Un'altra modalità di *déproverbialisation* consiste nel ridurre la cosiddetta estensione del proverbio. In effetti, sembra che una delle caratteristiche principali del proverbio sia l'estensione tendente all'assoluto, ossia la validità dello stesso in tutti i casi possibili. Questo è segnalato, dal punto di vista linguistico, dall'assenza di avverbi che esprimano i limiti dell'asserzione del proverbio o la sua validità assoluta. Inserendo nella formula proverbiale elementi in grado di ridurre la sua estensione o di esprimere riserve sulla sua qualità di verità generale, il parlante attiva il doppio meccanismo di azione di *proverbialisation* e *déproverbialisation* (Schapira, 2000: 94). In effetti, utilizzando tali forme avverbiali, il locutore altro non fa che ribadire il valore di verità assoluta del proverbio e, al contempo, segnalare il proprio disaccordo circa la validità dello stesso in certi casi specifici. Di seguito si riportano le traduzioni di alcuni degli esempi proposti dall'autrice:

- (9) a. L'appetito non viene sempre / necessariamente mangiando.
b. Ci sono notti che non portano consiglio.
c. Una rondine fa primavera?

Continuando la trattazione della *déproverbialisation*, non possiamo non parlare di *détournement* paremiologico, utilizzando la terminologia di Schapira (2000: 94-95). Per alcuni studiosi, tale fenomeno riguarda quegli enunciati che possiedono tutte le caratteristiche linguistiche del proverbio, ma che non fanno parte dell'insieme dei proverbi noti a una certa comunità. Secondo la studiosa, invece, si tratta piuttosto di una modifica eseguita nei confronti di un'espressione percepita come fissa per mezzo di una manovra lessicale, semantica o stilistica che crea un senso discorsivo a partire dal senso linguistico

dell'espressione originale. Per la studiosa, dunque, si tratta della deformazione di un proverbio esistente oppure di una creazione originale sulla base di uno stampo proverbiale che potrebbe essere immediatamente riconosciuto e identificato come tale. Anche per Michaux (1999: 90-91) si tratta di imitazioni di proverbi esistenti, oppure di modifiche di proverbi attestati che conservano alcuni elementi della formula iniziale cambiandone altri o ancora di frasi proverbiali create *ex novo* a partire da uno stampo proverbiale. In tutti questi casi di manipolazioni di proverbi, notiamo, su suggerimento di Michaux (1999: 91), che la rigidità formale dei proverbi è ciò che permette le modifiche dei proverbi stessi. Se i proverbi non fossero facilmente riconosciuti grazie alla loro fissità, allora non sarebbe nemmeno possibile operare quelle modifiche che certamente “deproverbializzano”, ma che rendono anche riconoscibile, il proverbio (o i proverbi) da cui trae origine il nuovo enunciato. Cambiando prospettiva, è possibile sostenere che senza la fissità non esisterebbe il processo di *déproverbialisation* che, invece, è così importante nella nostra società. Tuttavia, se da un lato la fissità è essenziale per la *déproverbialisation*, dall'altro lato notiamo che è proprio la *déproverbialisation* che spinge molti studiosi a intendere i proverbi come espressioni non fisse. Alla luce di queste considerazioni, è necessario sottolineare quest'ultimo passaggio: i proverbi possono subire manipolazioni oppure è possibile creare enunciati che ricordano proverbi esistenti ma tali possibilità non sono una prova della non fissità dei proverbi, bensì sono motivate dall'esistenza della loro fissità. Se i proverbi, infatti, non fossero espressioni fisse che i parlanti di una certa comunità in un dato momento conoscono e riconoscono, pur nelle loro diverse varianti, non sarebbe possibile identificare il proverbio originale che genera il proverbio creato o modificato.

1.6. Funzionamento dei proverbi nel discorso

Nei precedenti paragrafi è emersa una sostanziale divergenza di opinioni tra i diversi studiosi che più o meno recentemente hanno dedicato parte della loro ricerca all'analisi dei fenomeni sentenziosi, gnomici, paremiologici e idiomatici. Tutte le intuizioni emerse possono essere, da un lato, condivisibili e, dall'altro, discutibili. Questo perché si trovano molto spesso sostenute da esempi costruiti *ad hoc* che sembrano più rispondenti alle esigenze di ogni singolo ricercatore che alla realtà dell'enunciazione. Fatto, questo, messo in luce anche da Schapira (2000: 93) a proposito della negazione del messaggio proverbiale: la possibilità di negare un proverbio con un “no” categorico oppure attraverso l'accettazione del suo messaggio (con un “sì”, “certo”) e una rinegoziazione (“però”) viene supportato da vari

studiosi per mezzo di esempi creati *ad hoc*. Come sottolinea la studiosa, bisognerebbe studiare materiale autentico per vedere se tali pratiche vengono adottate e con quali modalità.

La *corpus linguistics*, assieme alla pragmatica, potrebbe essere di grande importanza nell'ambito di uno studio come quello che si sta delineando in queste pagine, data la grande complessità e arbitrarietà della materia trattata. Ecco perché è nostra ferma convinzione che il contesto d'uso e di enunciazione dei proverbi sia di fondamentale importanza per meglio comprenderne il funzionamento.

In questa sezione, ci si focalizzerà sul funzionamento dei proverbi nel discorso, per dare un inquadramento teorico generale alla pragmatica legata alla paremiologia.

Per prima cosa, si vedranno le principali modalità di inserimento dei proverbi nel discorso libero. Un numero cospicuo di ricercatori hanno rimarcato il fatto che al momento dell'utilizzo dei proverbi, i parlanti inseriscono formule volte a indicare che l'enunciato pronunciato o che si produrrà non è costruito liberamente, bensì è una sorta di citazione. Shapira (2000: 89-90) nota che in lingua francese si utilizzano spesso espressioni del tipo "*comme on dit*" o "*comme dit le proverbe*", sebbene siano frequenti i casi in cui i proverbi non sono introdotti da alcuna formula. Tale pratica evidenzia che, per la comunità linguistica, o quanto meno per gli attori coinvolti nella situazione comunicativa, la formula proverbiale è nota e il suo status di citazione (e la conseguente polifonia) è talmente evidente da rendere superflui ulteriori commenti.

Tuttavia, in letteratura la questione dell'inserimento dei proverbi nel discorso sembra più complicata. In primo luogo, Schapira (2000: 90), citando i lavori di altri autori come Michaux (1996), Anscombe (1989 e 1994) e Kleiber (1999), evidenzia che i proverbi non composizionali sono di più difficile inserimento nel contesto rispetto ai proverbi composizionali. Se pensiamo ai proverbi introdotti da verbi di opinione²⁸, notiamo che l'utilizzo di questi con quelli non composizionali è più discutibile rispetto all'impiego degli stessi con quelli composizionali. La ragione di tale diverso grado di integrazione all'interno del discorso libero è probabilmente dovuta al fatto che i proverbi non composizionali siano caratterizzati da una *proverbialisation* più spiccata e quindi da una maggiore stereotipia. Inoltre, i giudizi personali possono accompagnare i proverbi a condizione che si accetti la verità dello stereotipo espresso dal proverbio.

²⁸ Il che comporta, in realtà, il fenomeno di *déproverbialisation* precedentemente citato.

Nel contesto d'uso è probabile vedere i proverbi legati a deittici ed espressioni circostanziali come “qui”, “oggi”, “ieri”, “domani”, “in Francia”, ecc. Per accettarne l'uso, si prenderà in considerazione il fatto che anche in questo caso siamo di fronte a un fenomeno che rientra sotto l'ombrello della *déproverbialisation*, dato che si altera la condizione di genericità del proverbio, diventando, esso, un enunciato legato a una circostanza ben precisa.

Seguendo il ragionamento di Schapira (2000: 91-92), si nota che, se si passa al piano linguistico, solamente alcuni proverbi possono essere introdotti nel discorso attraverso verbi di opinione, senza ulteriori espressioni che segnalino il loro status. Ad esempio, i proverbi aventi un verbo alla forma imperativa non possono essere introdotti nel discorso senza altre etichette che mettano in luce la loro natura paremiologica e questo per chiare ragioni sintattiche: non si può, dunque, dire **Credo che aiutati che il ciel ti aiuta*. Se invece pensiamo ai proverbi contenenti vocaboli arcaici, vediamo che essi sono facilmente inseribili nel discorso senza alcuna introduzione, in quanto la loro natura citazionale è facilmente riconoscibile. L'uso di giudizi individuali indica semplicemente che nella situazione presa in esame si applica quanto descritto dal proverbio citato.

Dallo studio di Schapira emerge che di frequente i proverbi non sono accompagnati da alcuna espressione indicante il loro status. A volte, si possono trovare espressioni introduttive che non segnalano la loro natura di proverbio ma semplicemente il fatto che sono espressioni note ai parlanti di una comunità linguistica e che fanno parte della *doxa* della stessa comunità. Tali espressioni, che riprendiamo direttamente dall'articolo di Schapira (2000: 92), possono essere: “*On le sait*” e “*La bonne sagesse populaire*”. Vedremo nel seguito del nostro studio quali possono essere le espressioni equivalenti in altre lingue e, segnatamente, in italiano.

Relativamente all'uso dei proverbi nel discorso libero, è interessante notare la loro funzione argomentativa che molti studiosi definiscono *law-like* e che viene delineata in maniera chiara da Anscombe (1997: 46-49). Ciò significa, appunto, che il proverbio può essere impiegato nella stessa maniera in cui un avvocato utilizza la legge: il legale non ne è l'autore, ma le utilizza ai fini della propria argomentazione e per giungere alle proprie conclusioni. Allo stesso modo, il parlante che utilizza i proverbi non ne è l'autore e non li adopera nemmeno per descrivere una situazione. Semplicemente, tramite il proverbio si mira a qualificare una situazione, che diventa un caso particolare della situazione generica descritta dal proverbio.

Inoltre, il proverbio può inserirsi in un ragionamento come un'argomentazione a supporto di una tesi, ecco perché viene citato tale quale, senza essere introdotto da altre espressioni. Da questa considerazione deriva la non accettabilità del proverbio in posizione di conclusione dell'argomentazione. Per questa ragione, il proverbio combina con connettori del tipo "dato che", "posto che", "siccome", che ne mostrano il valore di argomentazione o premessa, ma non può essere introdotto da connettori come "pertanto", "per tale ragione", "in conclusione", ecc. In altri termini, non è accettabile un enunciato del tipo **Pertanto, una rondine non fa primavera.*

Tutte queste considerazioni legate alle modalità d'uso dei proverbi nel discorso libero sono di fondamentale importanza per considerazioni legate alla linguistica dei *corpora* in campo paremiologico, ovvero al reperimento automatico di proverbi e dei loro cotesti. Come si diceva, la *corpus linguistics*, tramite un procedimento di ricerca *bottom-up*, che parte dall'analisi di esempi reali per reperire ricorrenze e regolarità non influenzate da preconcetti, è in grado di studiare aspetti dei proverbi ancora ignoti o di convalidare teorie già emerse. È questo l'approccio che abbiamo deciso di seguire nel corso della presente Tesi di Dottorato.

2. Linguistica dei corpora e proverbi

2.1. Una scelta metodologica, ma non solo

Nello scorso capitolo, sono state delineate le basi teoriche essenziali per procedere all'analisi linguistica del proverbio, facendo non solo riferimento alla paremiologia in senso stretto, ma anche alla fraseologia, una disciplina che presenta numerosi punti di contatto, ma anche non poche divergenze con la paremiologia. Nel secondo capitolo affronteremo i principi teorici della metodologia da noi adottata per procedere all'analisi paremiologica del materiale linguistico di cui ci siamo avvalsi nel corso della presente ricerca, vale a dire, da un lato, la *linguistica dei corpora* anche detta *corpus linguistics*, dato il notevole sviluppo di tale metodologia in ambito anglosassone, e, dall'altro lato, l'approccio *corpus-driven*, frequentemente adottato nell'ambito di tale metodologia.

La linguistica dei corpora, in realtà, non è una semplice metodologia di analisi di fenomeni linguistici: è un vero e proprio approccio filosofico alla ricerca linguistica, che ha portato a grandi cambiamenti nei modi in cui si intende la lingua e nella qualità delle osservazioni prodotte sui fenomeni linguistici analizzati (Tognini-Bonelli, 2001: 1). Potremmo asserire che la linguistica dei corpora è al contempo un approccio alla ricerca linguistica, un campo di ricerca in sé e, dunque, una disciplina che è emersa, si è sviluppata e si è definita grazie alle tecnologie informatiche applicate all'analisi linguistica, nonché allo sviluppo dei primi corpora linguistici elettronici. Essendo, al contempo, una disciplina e un approccio alla ricerca linguistica, essa ha uno status teorico sufficientemente generale da essere applicato a una serie molto nutrita di campi di ricerca, quali la lessicografia, la didattica delle lingue, la traduzione, la stilistica, la grammatica, gli studi di genere, la linguistica computazionale, solo per citarne alcuni (cfr. Tognini-Bonelli, 2001: 1). Per questa ragione, e siccome tale metodologia ha già prodotto risultati soddisfacenti in ambito fraseologico, siamo convinti che, con i dovuti adattamenti, essa possa essere applicata alla paremiologia.

Sebbene la presente Tesi di Dottorato non voglia occuparsi della linguistica dei corpora *tout court*, bensì miri piuttosto a servirsene come strumento di analisi, è tuttavia necessario operare qualche considerazione di base in riferimento a tale metodologia. È lecito chiedersi perché si sia deciso di applicare allo studio dei proverbi la linguistica dei corpora: l'obiezione che si potrebbe muovere nei confronti di tale scelta è che non mancano di certo le risorse (studi pubblicati, dizionari, opere e raccolte di proverbi, *inter alia*) per analizzare i

proverbi in ottica culturale e interculturale, diacronica e sincronica, intralinguistica e interlinguistica. La risposta a tale obiezione è racchiusa nei principi fondamentali della linguistica dei corpora.

2.1.1. Nascita e sviluppo della linguistica dei corpora

Appare da subito molto interessante capire perché sia nata la linguistica dei corpora. Rispondere alla domanda “*Da quali bisogni è nata la linguistica dei corpora?*” equivale a dar conto di buona parte dei motivi che ci hanno spinto a scegliere questa metodologia per lo studio in oggetto.

Come si diceva, e in linea con quanto afferma Tognini-Bonelli (2001: 5), la linguistica dei corpora è nata con lo sviluppo delle nuove tecnologie che hanno permesso la riduzione dei tempi nell’analisi linguistica di quantità di informazioni che, prima dell’avvento delle stesse tecnologie, erano inimmaginabili. Il computer, inteso come strumento di analisi linguistica, ha certamente cambiato i parametri della ricerca, non solo perché ne ha accorciato i tempi, ma anche perché l’ha sistematizzata e ne ha permesso l’applicazione a quantità di dati sempre più importanti. “The computer has made possible unrestricted access to the observation of language in use” (*ibid.*) e il potere analitico del computer ha apportato innovazioni qualitative non trascurabili all’osservazione linguistica. Tuttavia, non si è arrivati alla linguistica dei corpora solamente perché da un certo momento in poi si è reso disponibile uno strumento (il computer) e materiali da analizzare (grandi corpora linguistici) prima inesistenti. Fin dagli esordi, la linguistica dei corpora si è proposta di analizzare dati reali per formulare ipotesi in grado di rendere conto dei fenomeni linguistici osservati tramite un procedimento induttivo. Tale obiettivo ne ha permesso, dunque, l’affermazione in ambito scientifico, assieme al conseguente sviluppo.

Si deve probabilmente a J. R. Firth (1890-1960) e alla sua *contextual theory of meaning* la nascita e lo sviluppo della linguistica dei corpora, nonché la sua applicazione alla ricerca linguistica, sebbene, in realtà, Firth sia deceduto prima dell’avvento dei corpora elettronici. Tognini-Bonelli (2001: 157) analizza ed espone la teoria di Firth che poggia principalmente su due considerazioni di base: da una parte, la linguistica descrittiva deve studiare il significato, d’altra parte lo stesso significato può essere espresso in termini linguistici. Come si vedrà, la linguistica dei corpora, e in particolare l’approccio *corpus-driven*, mira a descrivere il significato in senso lato partendo dalle sue realizzazioni a livello linguistico. In effetti, come illustra Tognini-Bonelli (2001: 157-158), rifacendosi ancora una

volta alle idee di Firth, si possono combinare le dimensioni linguistiche ed extralinguistiche prendendo come oggetto di studio l'uomo, inteso come individuo che agisce nella realtà che lo circonda. Se consideriamo l'uomo come animale sociale e se accettiamo l'idea secondo la quale ciò che siamo, ciò che diciamo e ciò che facciamo sono intrinsecamente connessi, capiamo come qualsiasi attività umana abbia un significato. Ecco perché tutto quanto faccia parte di qualunque situazione ha un significato: i partecipanti sono significativi, così come gli oggetti inanimati e gli eventi. Da questa considerazione ne scaturisce un'altra: la capacità di comunicare non può esistere in completo isolamento e in assenza di contesto. Per la medesima ragione, il significato non può essere analizzato come un elemento isolato e senza prendere in considerazione l'evento comunicativo in cui occorre. Per questa ragione il fulcro della *contextual theory of meaning* proposta da Firth è, appunto, il contesto situazionale, in cui tutto è rilevante ai fini della costruzione del significato (cfr. Tognini-Bonelli, 2001: 158-19). Come sottolinea Tognini-Bonelli (2001: 157), nell'ambito della linguistica dei corpora si prendono in considerazione i dati derivanti dai corpora per disambiguare il significato di parole o espressioni. Ciò è possibile solamente perché, come già aveva notato Firth, il significato trova anzitutto la sua realizzazione a livello linguistico.

Un altro concetto fondamentale per la linguistica dei corpora è quello di “eventi ripetuti” (*ibid.*: 159-160). La lingua è un vettore della continuità di ripetizioni nei processi sociali: l'uomo, da un punto di vista linguistico, agisce in maniera sistematica. Da tale concezione emerge l'importanza di identificare pattern ricorrenti²⁹ per analizzarli e sistematizzarli tramite una serie di categorie descrittive. Per fare ciò, diventano di estrema importanza i corpora linguistici, che permettono al linguista di quantificare tale continuità di ripetizione che sarà necessaria per produrre considerazioni tipologiche. Da un lato, il parlante deve essere considerato sia come parlante individuale sia come utente tipo della lingua. Dall'altro lato, il linguista deve riuscire ad astrarre l'impersonale, ossia il ricorrente, dal personale e a considerare il primo come tipologico. L'unica maniera che abbiamo per distinguere fatti tipologici da fatti individuali è andare alla ricerca di eventi ripetuti. Solo individuando fatti ricorrenti è possibile arrivare a capire che questi sono impersonali e tipologici e non semplicemente dettati da caratteristiche personali o dalla casualità della contingenza. La lingua, infatti, viene da un lato impiegata per perseguire un effetto in un determinato contesto e dall'altro lato è costituita da una serie di elementi abituali e ricorrenti

²⁹ Per una definizione di *pattern ricorrenti*, si veda § 2.7.

che agiscono come un criterio di rilevanza: in altri termini, più un certo elemento è ricorrente e più è rilevante ai fini linguistici. Ecco perché la linguistica, il cui oggetto di studio è ciò che è “tipico”, non può prescindere dall’uso reale (Tognini-Bonelli, 2001: 160).

2.1.2. Corpora linguistici

Prima di proseguire, è opportuno gettare luce sulla nozione di corpus linguistico, specialmente per proporre una definizione nell’ambito della linguistica dei corpora. Tali considerazioni saranno funzionali ai § 2.3 e 2.4, in cui si tenterà di descrivere i corpora utilizzati nel corso del presente studio.

Nel terzo capitolo della sua monografia, Tognini-Bonelli (2001: 52-55) propone diverse definizioni di corpus elaborate da diversi studiosi, quasi a indicare le molteplici sfaccettature, nonché i diversi obiettivi, dei corpora. Si è generalmente concordi nell’asserire che un corpus sia una raccolta testuale, anche se non si può affermare che nei corpora ci siano esclusivamente testi integrali. In base alle diverse definizioni presentate e all’esperienza maturata sul campo, possiamo affermare che è fuori dubbio ritenere che i corpora siano raccolte di campioni di testi, il che implica che la lingua che vi si ritrova è una lingua effettivamente in uso nel momento in cui sono stati prodotti i testi o gli stralci di testi presenti nei corpora.

Non a caso è stato impiegato il termine “campione” per fare riferimento ai testi che troviamo nei corpora. In ambito scientifico, la campionatura è a sua volta determinata da criteri di selezione volti a raggiungere un determinato obiettivo. Nel caso dei corpora, gli obiettivi da raggiungere possono essere prettamente linguistici oppure di altra natura, come si vedrà più avanti in questo § 2.1.2, quando si parlerà più dettagliatamente della finalità dei corpora. Quest’ultima è una variabile da tenere in considerazione sia nella fase di costruzione dei corpora sia nella fase della loro consultazione.

Generalmente, nei corpora devono essere presenti tre caratteristiche di base che devono essere tenute in considerazione tanto dai loro autori quanto dagli utenti finali. La prima è l’autenticità, la seconda è la rappresentatività e la terza è la finalità.

L’autenticità del materiale raccolto in un qualsiasi corpus è al contempo il primo criterio che gli autori del corpus devono prendere in esame e l’assunto di base che permette all’utente del corpus di produrre considerazioni linguistiche di qualsiasi natura. Autenticità significa che la lingua presente nei testi che compongono i corpora sia quella effettivamente utilizzata in situazioni ordinarie dai parlanti della stessa comunità linguistica (Tognini-Bonelli,

2001: 55). In altri termini, il materiale del corpus non può essere prodotto ad hoc, ma deve essere preso da fonti esterne e aggiunto al corpus stesso senza subire alcuna modifica.

Un discorso a parte, invece, deve essere fatto per la *rappresentatività* dei corpora. Prima di illustrare le diverse sfaccettature di questa caratteristica, consideriamo il suo significato in statistica, disciplina da cui è stato preso in prestito. La rappresentatività è legata ad altri due concetti le cui etichette sono *popolazione* e, di nuovo, *campione*. “Una *popolazione* è un insieme di tutte le possibili osservazioni di un tipo su un dato campo. Un *campione* invece è una sezione, una parte della popolazione, che include solo alcune delle possibili osservazioni” (Chiari, 2007: 42). La rappresentatività entra in gioco quando si mettono in relazione una popolazione con il suo campione: un campione, infatti, per essere utile e significativo deve essere rappresentativo di una popolazione, vale a dire “essere atto a esibire lo stesso tipo di informazioni (qualitativa) con la stessa probabilità di occorrenza (quantitativa) della popolazione” (*ibid.*: 43). Ecco che si delineano le principali caratteristiche della rappresentatività anche nella linguistica dei corpora, nonché le principali problematiche legate alla costruzione dei corpora. Se i campioni devono essere in un rapporto di omologia con la relativa popolazione, vale a dire rispecchiare nella maniera più fedele possibile una popolazione, si deve porre una grande cura nella scelta dei campioni stessi, affinché siano ben bilanciati (rispecchino dunque la diversa distribuzione qualitativa della popolazione nelle medesime proporzioni) e in grado di restituire le stesse informazioni che si potrebbero reperire presso la popolazione rappresentata. Si capirà fin da ora che, nonostante tutte le cure necessarie riposte nella formazione di un corpus e malgrado gli enormi passi avanti effettuati in questo ambito in linguistica dei corpora, è davvero difficile (se non impossibile) calcolare il grado di rappresentatività di un corpus nei confronti di una certa popolazione, nonché arrivare a comporre corpora davvero rappresentativi. Ciononostante, quando si utilizzano corpora linguistici è possibile delineare tendenze di massima che potrebbero essere valide, nella stessa misura e per la medesima popolazione, senza la pretesa di giungere a conclusioni assolutamente certe, ma con la convinzione di produrre teorie statisticamente valide e il più possibile prossime alla realtà delle cose.

Soffermiamoci ora sul pensiero di alcuni autori in relazione alla rappresentatività dei corpora, così come riportato in Tognini-Bonelli (2001: 57). Secondo Leech (1991), un corpus è rappresentativo quando le conclusioni a cui porta possono essere generalizzate e valide per corpora più ampi, mentre Biber (1994) sposta il punto di vista e afferma che la maggior parte

degli usi di un corpus per finalità di ricerca linguistica si basano sulla convinzione che il corpus sia uno strumento rappresentativo della lingua. In altri termini, l'assunto da cui si parte prima di utilizzare un corpus per produrre analisi linguistiche è che questo sia rappresentativo di una certa popolazione e che i risultati che emergeranno siano applicabili a un campione più ampio oppure alla lingua nel suo insieme. Tuttavia, la rappresentatività del corpus è anche un atto di fede (Leech, 1991), dato che non vi sono mezzi per garantire una totale rappresentatività del corpus o per valutare tale dimensione in maniera obiettiva. Ciononostante, la rappresentatività del corpus è di grande importanza non soltanto per l'utente del corpus stesso, ma anche per il suo autore. Biber (1994) ritiene che sia necessario tenere in considerazione la rappresentatività nella fase di composizione del corpus. A tal proposito, egli considera che la rappresentatività sia la capacità di un campione di includere l'intera gamma di variabilità esistente in una popolazione, non soltanto dal punto di vista situazionale ma anche da quello linguistico. Biber (1993: 244-245) precisa che il genere e il registro rientrano nelle categorie testuali in un'ottica situazionale, mentre i tipi testuali sono categorie definite da un punto di vista linguistico, ovvero, diversamente dai generi e dai registri, sulla base di pattern linguistici co-occorrenti condivisi. Ciò significa, inoltre, che i generi e i registri dipendono da criteri esterni al corpus stesso, mentre i tipi testuali si basano su fattori interni. Ad esempio, i registri sono legati a fattori quali le situazioni, gli scopi e le funzioni assunti da un testo all'interno di una comunità linguistica e tali criteri possono essere individuati in una fase precedente alla costruzione del corpus. Al contrario, non vi sono mezzi per individuare a priori i tipi testuali, in quanto questi dipendono da caratteristiche linguistiche interne al corpus analizzabili grazie a un corpus rappresentativo di testi.

La terza caratteristica dei corpora è la loro *finalità*. Gli utenti di un corpus (che potrebbero essere ricercatori, linguisti, professori e discenti di lingue straniere) utilizzano un determinato corpus in virtù della sua natura (il fatto di essere un bacino di prove) per eseguire delle analisi di carattere linguistico. Corpora generali utilizzati per questa finalità sono definiti *corpora di riferimento*, in quanto mirano a essere rappresentativi della lingua stessa (cfr. Chiari, 2007: 51 e § 2.1.3). Il discorso cambia se gli obiettivi sono altri e se perdono questo carattere di genericità: ad esempio, se lo scopo è analizzare un corpus di traduzioni della bibbia oppure studiare un corpus di interviste con pazienti psichiatrici. In questi casi, e specialmente se il contesto comunicativo dei testi che entrano a fare parte dei corpora è volutamente artificiale o sperimentale, non si è più di fronte a un corpus di riferimento ma ad

altri tipi di corpora specifici per determinati ambiti. Solitamente, gli autori di questi corpora forniscono da subito precisazioni sulla diversa natura dei corpora, in maniera da evitare usi errati degli stessi, che non possono essere utilizzati per studi linguistici di carattere generale. Possiamo anticipare, fin da ora, che nell'ambito della presente ricerca di Dottorato si farà uso di corpora di riferimento o aventi caratteristiche simili a questi.

2.1.3. Tipi di corpora

Si intuirà, a questo punto, che esistono diversi tipi di corpora: in effetti, e a ragione, in linguistica dei corpora si parla generalmente di *corpora* al plurale, essendoci una differenza molto accentuata tra loro, sia per tipologia sia per scopi perseguiti. Per le medesime ragioni, il ricercatore utilizzerà certi tipi di corpora e non altri in funzione della tipologia di risultati che vuole ottenere o del tipo di ricerca che mira a condurre.

Cominciamo questa breve carrellata assumendo una prospettiva intralinguistica. Ad esempio, per alcuni studi si può rendere necessario un corpus volto a “registrare tutte le principali varietà di una lingua, e quindi essere rappresentativo della lingua stessa” (Chiari, 2007: 51). Come si è già visto, un corpus di questo tipo viene definito solitamente *corpus di riferimento* (o *reference corpus* in ambito anglosassone). Simili corpora, solitamente di grandi dimensioni, mirano a raccogliere testi appartenenti a diverse varietà sociolinguistiche, diafasiche e diatopiche e a rappresentare in maniera bilanciata diverse tipologie testuali (Chiari, 2007: ibid.). Come evidenzia Chiari (2007: ibid.), lo “standard fino agli anni Novanta si aggirava intorno a circa 500.000 parole, oggi, con la disponibilità di maggiori risorse tecnologiche e con il web come deposito immenso di materiali, i corpora di riferimento più diffusi possiedono un'estensione di circa 100 milioni di parole”. Un'affermazione recente se pensiamo all'anno di pubblicazione della monografia di Chiari (2007) ma oramai superata se consideriamo che corpora oggi esistenti hanno già superato tale soglia. Similmente ad altri ambiti tecnologici o informatici, anche nel caso della linguistica dei corpora si compiono enormi passi avanti con ritmi estremamente incalzanti.

In questa breve carrellata, accenneremo soltanto alla questione dell'utilizzo dei corpora per l'analisi della terminologia di certi domini, senza entrare nel dettaglio di questi corpora, in quanto essi rivestono uno scarso interesse ai fini della presente ricerca. Si nota, infatti, che tra le possibilità offerte dalla linguistica dei corpora all'analisi della lingua, vi è quella di restringere il campo a certe varietà linguistiche. Ad esempio, si potrebbero costruire corpora ad hoc utilizzando articoli di quotidiani economici per una prima estrazione

terminologica in ambito economico oppure per introdurre studenti di economia al linguaggio tecnico della loro area (Tognini-Bonelli, 2001: 8). Notiamo, dunque, come i corpora siano di grande importanza per lo studio di terminologie specifiche, non solo per quanto attiene all'individuazione della terminologia *tout court*, ma anche per l'identificazione del significato attraverso la formalizzazione dei pattern contestuali, i quali si rivelano altresì di grande utilità per i discenti nella formulazione delle definizioni terminologiche (*ibid.*).

Anzitutto, se ci si concentra sulla sfera delle risorse multilingue, notiamo che vi è una notevole differenza tra i cosiddetti *corpora paralleli*³⁰ e i *corpora comparabili*³¹. Entrambi sono corpora multilingui, ovvero formati da testi in almeno due lingue. La differenza, tuttavia, risiede nel fatto che i primi sono corpora composti da testi in lingua di partenza e dalle traduzioni degli stessi, mentre i secondi sono certamente corpora multilingui ma non sono composti da testi fonte e dalle relative traduzioni. Semplicemente, questi ultimi corpora sono formati da campioni di testi simili in lingue diverse scelti in base a certi criteri esterni che devono essere rispettati dai testi delle lingue presenti nel corpus, come il registro, il canale (ad esempio lingua scritta / lingua orale), ecc.

Se, dunque, i primi possono essere allineati, i corpora comparabili possono solamente essere confrontati per cercare delle corrispondenze che tuttavia non sono il frutto di un processo traduttivo. Come si vedrà anche nel corso della presente ricerca, le conclusioni a cui permettono di giungere i due tipi di corpora sono notevolmente diverse. Anticipiamo, anche in questo caso, che nell'ambito della presente ricerca di Dottorato, si è scelto di lavorare con corpora comparabili anziché corpora paralleli in ragione delle finalità da perseguire.

2.2. L'analisi dei proverbi in corpora linguistici e altri approcci

Prima di illustrare i corpora impiegati e l'approccio adottato nel corso di questa tesi, è bene soffermarsi su qualche considerazione fondamentale circa i motivi che ci hanno spinto a studiare i proverbi in corpora e non semplicemente in compendi, liste, dizionari di proverbi, opere letterarie più o meno recenti o opere generalmente riguardanti i proverbi. Le ragioni di tale scelta sono direttamente legate al punto di vista adottato nell'ambito di questa ricerca.

³⁰ Sintagma ripreso da Chiari (2007: 53-54).

³¹ Chiari (2007: 53-54) li definisce *corpora multilingui*. Per evitare confusione con quanto noi crediamo sia la categoria dei corpora multilingui, ossia la macrocategoria che racchiude al suo interno i corpora paralleli e quelli comparabili, preferiamo impiegare, traducendolo in italiano, il sintagma inglese *comparable corpora* (cfr. Tognini-Bonelli, 2001: 6-7).

La presente ricerca parte da una necessità concreta emersa in ambienti ben precisi: quelli della traduzione e dell'interpretazione. Più precisamente, questa ricerca si propone di iniziare un percorso verso la formulazione di risposte a diversi quesiti: come si devono comportare traduttori e interpreti quando si trovano di fronte a un proverbio? Quali sono le caratteristiche del proverbio che devono essere rispettate e quali invece possono essere modificate e adattate ai testi di arrivo? Come si possono reperire proverbi in lingua di arrivo ignorandone il significato compositivo ma conoscendone a priori il significato paremiologico? Come non rischiare di utilizzare in lingua di arrivo proverbi inesistenti, desueti, oppure dal significato paremiologico diverso rispetto ai relativi proverbi in lingua di partenza?

Questi sono solo alcuni dei quesiti che ci siamo posti in base a esperienze concrete sul campo, maturate sia in ambito accademico sia in ambito professionale. Come si diceva, non vi è la pretesa di giungere a una risposta definitiva per ognuno di questi interrogativi, bensì la volontà di fare chiarezza almeno sui punti principali che li accomunano. Di grande importanza, per l'intera ricerca, sarà il tentativo di gettare luce sulle caratteristiche maggiormente legate alla pragmatica del proverbio, in italiano e in francese, le due lingue scelte per questo studio. Pertanto, ci interesseremo a elementi quali la frequenza d'uso dei proverbi nelle diverse lingue, in maniera da indicare a un potenziale professionista delle lingue il bacino paremiologico entro cui cercare il proverbio in lingua di arrivo, ma anche alla maniera in cui vengono impiegati i proverbi (sono maggiormente utilizzati secondo la variante standard che troviamo nei dizionari oppure in versioni modificate?) e agli scopi che i parlanti intendono perseguire enunciando certi proverbi in determinate situazioni.

Per cercare di inquadrare meglio il nostro approccio, cerchiamo anzitutto di fare un confronto tra la metodologia da noi scelta per stilare liste di frequenza d'uso dei proverbi e quanto è già stato fatto in questo senso da ricercatori di diversa provenienza, specialmente nell'ambito della realizzazione dei cosiddetti *minimi paremiologici*. Tra i primi ideatori del concetto di minimo paremiologico figura il linguista e folklorista russo Grigoriĭ Permiakov (1919-1983) che compose una lista di 300 tra espressioni fisse, proverbi e altri modi di dire che secondo lui erano essenziali per avere una buona conoscenza della lingua russa. In epoca più recente, nell'ambito del minimo paremiologico, sono degni di nota, tra gli altri, i lavori del Gruppo di Ricerca UCM 930235 *Frasesología y Paremiología*, un team di ricercatori diretto da María Teresa Zurdo Ruiz-Ayúcar e Julia Sevilla Muñoz (Sevilla Muñoz, 2011: 77).

Sempre legata a questo gruppo di ricerca, ma impegnata tra l'altro nell'individuazione del minimo paremiologico italiano, ricordiamo Maria Antonella Sardelli (2011: 87-95). L'obiettivo del minimo paremiologico in una data lingua (rimarchiamo, infatti, che non può esistere un minimo paremiologico mondiale, in quanto i proverbi sono troppo legati alle specificità delle diverse culture in cui si sono sviluppati da consentire un'esatta corrispondenza presso popoli diversi) consiste nell'individuare l'insieme degli "enunciados sentenciosos estables más conocidos por una mayoría significativa de los hablantes de una comunidad sociocultural concreta" (Sevilla Muñoz, 2011: 77). Quanto alla metodologia adottata, presenteremo, in maniera tutt'altro che esaustiva, le principali tappe toccate nell'elaborazione del minimo paremiologico spagnolo e presentate in Sevilla Muñoz (2011).

Solitamente si procede a una prima scrematura delle cosiddette paremie individuando quelle più frequenti in documenti scritti, come possono essere raccolte di proverbi o opere generalmente relative ai proverbi. Una volta individuate, esse vengono passate al vaglio di indagini realizzate mediante la somministrazione di diversi questionari a campioni di utenti scelti a priori e suddivisi in diversi gruppi. Nell'ambito dell'esperienza spagnola, ad esempio, i campioni venivano scelti perlopiù in ambito accademico tra studenti di varia nazionalità, per ragioni legate alla formulazione di minimi paremiologici in altre lingue, e docenti, nonché nelle cerchie di amici e parenti degli autori dello studio. Lo scopo dei sondaggi era di comprendere il grado di conoscenza delle paremie individuate nonché il loro uso. Le paremie entravano a far parte del minimo paremiologico spagnolo solo se apparivano almeno nel 60% dei questionari compilati. L'informazione circa il loro uso, invece, dava accesso alla loro frequenza d'uso specialmente per un interesse legato all'insegnamento delle lingue straniere a seconda del livello di lingua da raggiungere.

Ciò che ci siamo proposti di compiere in questa sede è un lavoro per certi aspetti simile ma anche abbastanza diverso, il che ci impone di staccarci nettamente da qualsiasi tema legato al minimo paremiologico. Anzitutto, rispetto al minimo paremiologico, l'oggetto di studio è esclusivamente il proverbio secondo i criteri che sono stati stabiliti nel primo capitolo della tesi in oggetto e non una categoria generale di enunciati completamente o parzialmente sentenziosi e cristallizzati che vada dalle espressioni idiomatiche ai proverbi. In secondo luogo, la metodologia scelta è profondamente diversa. I proverbi che saranno oggetto di studio della nostra ricerca non verranno scelti da alcuna opera esistente (sebbene, come mostreremo in § 2.9.1, abbiamo esplorato anche questa strada, per poi abbandonarla). Una

delle ragioni per cui non utilizzeremo alcuna risorsa esistente (come dizionari o elenchi di proverbi) dipende dal fatto che tali opere vanno in direzione contraria alla scrematura dei proverbi in base al loro reale e contemporaneo utilizzo; esse puntano maggiormente sul massimo paremiologico (cfr. a tal proposito Sevilla Muñoz, 2011: 82) piuttosto che sul minimo paremiologico. I proverbi saranno invece individuati in corpora linguistici (secondo le modalità che verranno successivamente indicate). In terzo luogo, si evita in questa sede l'utilizzo di informatori per ricevere informazioni circa il grado di conoscenza e d'uso dei diversi proverbi. Abbiamo optato per questa scelta non solo perché, come sottolineato anche da Sevilla Muñoz (2011: 81), non è sempre facile trovare soggetti con determinate caratteristiche socioculturali disposti a collaborare a un simile progetto, ma anche e soprattutto perché era nostra ferma volontà analizzare quanto accade in contesti non artificiali, spontanei e autentici quando si utilizzano i proverbi. Al contrario, la logica del questionario da somministrare a certi soggetti non rispetta tale criterio di autenticità, per noi essenziale per elaborare uno studio davvero oggettivo, in grado di fornire risposte concrete ad alcune delle necessità dei professionisti delle lingue. Per tale ragione, una delle nostre priorità consisterà nell'eseguire un'indagine oggettiva, autentica e realistica sui proverbi, una ricerca che faccia "parlare" i dati autonomamente, affinché si possa giungere alla formulazione di determinate ipotesi esclusivamente sulla base di ricorrenze notate nell'uso ordinario dei proverbi. Questa volontà ricalca il principio fondamentale che ha dato vita all'intera *corpus linguistics*, che si è venuta ben presto a opporre alla cosiddetta *armchair linguistics*. Probabilmente questa sorta di dicotomia è indicativa del rapporto che intercorre tra il presente lavoro e i tradizionali approcci allo studio della paremiologia.

2.3. Caratteristiche dei corpora WaCky. Focus sul corpus itWaC

Il corpus iniziale che si è scelto di impiegare per questa ricerca di dottorato è il corpus chiamato itWaC. Esso fa parte di una raccolta di corpora linguistici, detti corpora WaCky (acronimo di Web as Corpus kool yinitiative), che condividono alcune caratteristiche fondamentali: essere corpora di grandi dimensioni, vale a dire costituiti da più di un miliardo di parole; essere formati da testi reperiti automaticamente sul Web, taggati per parti del discorso e lemmatizzati; servire a scopi di analisi delle lingue che rappresentano. Le principali informazioni riguardanti i corpora e le metodologie seguite per la loro costruzione sono presentate in Baroni et al. (2008): gli autori di questo articolo, ossia M. Baroni, S. Bernardini,

A. Ferraresi ed E. Zanchetta, sono tra gli artefici di alcuni dei corpora in oggetto e tra i principali esperti e promotori del progetto WaCky.

I primi corpora sono stati sviluppati in lingua inglese, scegliendo per quanto possibile la varietà britannica. In seguito ne sono stati aggiunti altri in lingua tedesca e italiana. I corpora³² così ottenuti sono stati denominati come segue: ukWaC, per la varietà britannica dell'inglese; deWaC, per il tedesco; itWaC, per l'italiano. In una fase successiva, è stata compiuta un'altra esperienza molto simile di generazione di corpora mediante ricerca automatica sul Web in lingua francese che ha dato vita al corpus frWaC. È stato, poi, aggiornato il sistema di annotazione linguistica del corpus ukWaC (che era solo *pos-tagged* e lemmatizzato), dando vita a PukWaC (aggiungendo al *pos-tagging* e alla lemmatizzazione il *parsing*), ed è stato costruito il corpus WaCkypedia_EN, una sorta di backup del Wikipedia inglese del 2009 avente lo stesso tipo di annotazione di PukWaC. Infine, è stato aggiornato il corpus deWaC in termini di eliminazione di alcuni casi di cosiddetto “rumore”, quali ripetizioni di vario genere. Il corpus così ottenuto è stato denominato SDeWaC.

Vediamo ora, nel dettaglio, i principali vantaggi di questi corpora, concentrandoci in maniera particolare sul corpus itWaC, impiegato nel corso della Ricerca di Dottorato in oggetto. L'utilità dei corpora WaCky risiede nella loro capacità di combinare una grande quantità di dati con la grande varietà e praticità di utilizzo dei corpora stessi, per fini linguistici non definiti a priori e potenzialmente illimitati, grazie alla loro lemmatizzazione e al *POS Tagging*. In altri termini, non solo si tratta di un'enorme quantità di dati a disposizione del ricercatore, ma anche di dati annotati per parti del discorso che quindi il ricercatore è in grado di sfruttare in maniera più efficace per perseguire i propri fini di ricerca linguistica nella più ampia libertà.

È bene, tuttavia, soffermarsi sulla tipologia dei dati. Di seguito, si esporrà il principio guida che ha ispirato la costruzione dei diversi corpora nelle varie lingue. L'obiettivo era ritrovare una serie di URL, ovvero di indirizzi di pagine web, in grado di essere rappresentativo di una nutrita varietà di contenuti e generi. Per ritrovare siti rispondenti a queste caratteristiche, coppie di parole lessicali sono state inserite in un motore di ricerca commerciale utilizzando un servizio API. Le coppie di parole lessicali sono state scelte casualmente tra le parole di frequenza media reperite in diversi modi a seconda delle lingue.

³² L'elenco completo dei corpora WaCky si trova alla seguente pagina web: <http://wacky.sslmit.unibo.it/doku.php?id=corpora>.

Da questa scelta, sono state da subito escluse le parole funzione. Ogni coppia di parole ha generato numerosi URL. Solamente dieci URL per ogni coppia di parole sono stati presi in considerazione, selezionandoli casualmente e ponendoli in un'unica lista. I duplicati di URL sono stati eliminati e per ogni dominio è stato mantenuto un solo URL selezionato a caso. La lista ottenuta e ripulita è stata, successivamente, inserita in un *crawler*,³³ software preposto all'analisi automatica dei contenuti di una rete, generalmente sulla base di una lista di URL che il programma andrà a visitare automaticamente. Le ricerche del software sono state limitate ai domini web rilevanti per ogni lingua (.it per l'italiano; .de e .at per il tedesco; .uk per l'inglese), escludendo gli URL che terminavano con un suffisso indicante una tipologia di dati non HTML (come .pdf, .jpg, ecc.).

I diversi corpora ottenuti erano per così dire grezzi. Gli autori hanno pertanto cercato di renderli versatili e fruibili per il più ampio spettro di utenti e di ricerche linguistiche, aventi scopi diversi ma basate sugli stessi corpora. Per questa ragione, si è resa necessaria una fase di pulitura e di annotazione dei dati.

Anzitutto, sono stati mantenuti esclusivamente file di testo di tipo html con dimensione tra 5 e 200KB. Questo perché è stato dimostrato che documenti troppo piccoli possono contenere pochi testi fruibili mentre documenti esageratamente grandi potrebbero contenere a loro volta liste di vario genere, di scarso interesse ai fini linguistici.

La parte più interessante della pulitura ha certamente riguardato la rimozione di duplicati di testi o di porzioni di testi ripetuti in documenti diversi. Dopo aver eliminato i documenti ripetuti, sono state dapprima individuate, e successivamente rimosse, quelle porzioni di testo ripetute in siti diversi. Si tratta di quegli elementi che tendono a essere uguali in tutti i documenti web (come intestazioni, piè di pagina, barre di navigazione, *disclaimer*, ecc.) e che rivestono uno scarso interesse linguistico, specialmente perché a una loro alta frequenza non corrisponde un'elevata rilevanza. Data l'enorme quantità di dati da analizzare, l'individuazione e la rimozione di questi elementi sono state automatiche, il che ha sicuramente garantito tempistiche molto ridotte e un buon grado di affidabilità, sebbene non si possa garantire l'eliminazione totale di tali elementi³⁴.

³³ Il software utilizzato è stato Heritrix. Si veda: <http://crawler.archive.org/>.

³⁴ Si vedrà in § 3.2.6 come effettivamente ci siamo accorti noi stessi, lavorando con i corpora itWaC e frWaC, della persistenza di ripetizioni all'interno dei corpora. Questo perché il procedimento suesposto riesce a eliminare solo i documenti identici, ma in rete esistono molti documenti quasi identici (che differiscono, ad esempio, per un link o un'immagine) impossibili da eliminare automaticamente.

In seguito a questa tappa, i testi rimanenti sono stati filtrati sulla base di liste di parole funzione per eliminare quei testi che rispettavano criteri minimi di coesione. Generalmente, i testi coesi contengono in proporzione molte parole funzione, pertanto i testi che non rispondevano a criteri testuali minimi, come “ten types and thirty tokens per page, with function words for at least a quarter of all words” (Baroni et al., 2008: 7), sono stati scartati. Con il medesimo filtro, in aggiunta, si sono potuti scartare quei documenti in lingua diversa da quella desiderata. Infine, dopo aver creato liste di parole frequentemente utilizzate in siti pornografici, sono stati eliminati questi tipi di risultati, che contengono testi anche di grandi dimensioni generati artificialmente in maniera automatica, probabilmente per ingannare i motori di ricerca.

La seconda fase di pulitura dei corpora si è concentrata sulla rimozione di quelle porzioni di testo identiche reperite in documenti diversi. Per fare ciò, sono state selezionate casualmente combinazioni di parole, dette *n-gram*, senza contare le parole funzione. Successivamente, si sono confrontati gli *n-gram* dei vari documenti. Quando una quantità *t*, decisa a priori, di *n-gram* ricorreva in più documenti, questi erano eliminati. In questa fase di ulteriore pulitura dei testi, i gruppi di parole confrontati erano di cinque parole, detti *5-gram*, e per ogni testo venivano selezionati casualmente 25 *5-gram*. Si procedeva all’eliminazione di un documento se questo conteneva almeno due *5-gram* uguali ad almeno un altro documento.

Queste due fasi di pulitura dei testi hanno prodotto risultati molto interessanti. Da notare, infatti, la drastica riduzione delle dimensioni dei corpora dopo tali processi. In media, le dimensioni finali dei corpora si erano ridotte del 96% rispetto ai relativi corpora grezzi. Si è trattato, dunque, di un enorme processo di pulitura, nonostante il quale le dimensioni finali dei corpora sono rimaste di grande rilevanza.

Conclusa la fase di pulitura dei corpora, gli autori hanno proceduto all’annotazione degli stessi per parti del discorso. Per la lemmatizzazione dei corpora ukWaC, deWaC e frWaC è stato usato il software TreeTagger³⁵, mentre per il corpus itWaC è stato usato il software Morph-it!³⁶.

2.4. Perché itWaC e frWaC

Le ragioni per cui si è deciso di affrontare lo studio dei proverbi sui corpora itWaC e frWaC sono molteplici. Da un lato, vi sono motivazioni quantitative e qualitative riguardanti

³⁵ URL: <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>.

³⁶ URL: <http://dev.sslmit.unibo.it/linguistics/morph-it.php>.

tali corpora, dall'altro vi sono ragioni pratiche legate alla tipologia di studio dei proverbi e alla finalità da perseguire.

Ci si concentrerà dapprima sulla validità di operare con i corpora WaCky per raggiungere le finalità di ricerca connesse al presente studio. Si tenterà, dunque, di individuare i punti di forza e i punti deboli di tali corpora, facendo particolare riferimento al corpus itWaC, non solo perché sarà uno dei corpora che utilizzeremo, ma anche per la maggiore disponibilità dei dati che lo riguardano³⁷. Ad ogni modo, si tenga conto che le sue caratteristiche valgono anche per gli altri corpora WaCky, e quindi anche per il corpus frWaC, dato che i processi che hanno portato alla loro realizzazione sono i medesimi.

In primis, si confronterà itWaC con il corpus de *la Repubblica*. A titolo informativo, questo corpus raccoglie tutti i numeri pubblicati in 16 anni. Si tratta di un corpus costituito interamente da articoli di giornale, ma data la varietà degli argomenti affrontati e la quantità di dati che è in grado di mettere a disposizione è un riferimento di grande importanza nel panorama linguistico italiano.

Il confronto più immediato che si può realizzare tra il corpus de *la Repubblica* e itWaC è di tipo quantitativo. *In primis*, notiamo che il corpus itWaC è molto più ampio di quello de *la Repubblica*: se nel primo ci sono 941.990 *type* appartenenti alla categoria dei sostantivi, nel secondo ce ne sono solamente 218.893; il primo contiene 706.330 *type* appartenenti alla categoria degli aggettivi, il secondo solamente 145.300; infine, se itWaC contiene 679.758 *type* verbali, *la Repubblica* ne contiene soltanto 140.342. Concentrandoci sui *type* aventi almeno 20 *token* ciascuno, che sono più significativi di quelli aventi una frequenza inferiore, considerando che la soglia di ricorrenza minima necessaria a formulare una descrizione lessicografica di una parola è di 20 volte (cfr. Sinclair, 2005), si nota che il numero di questi *type* più significativi è comunque maggiore nel corpus itWaC che nel corpus de *la Repubblica* (al netto del rumore, come errori di battitura, testi sconnessi e forestierismi, calcolato statisticamente). Potrebbe sembrare un'ovvietà, in quanto itWaC supera il corpus de *la Repubblica* per numero di *type* e di *token*. Tuttavia, è significativo, in quanto è un'ulteriore prova a supporto della validità e dell'affidabilità di itWaC, che viene a essere un corpus in grado di rispondere a un maggior numero di interrogazioni e di poter fornire informazioni utili anche su fenomeni generalmente poco frequenti in lingua, come parole aventi una bassa frequenza d'uso (Baroni et al., 2008: 14).

³⁷ Cfr. Baroni et al. (2008).

La dimostrazione che i corpora WaCky, oltre a raccogliere una grande quantità di dati, sono anche affidabili è di fondamentale importanza ai fini della scelta del corpus per la presente ricerca. Sappiamo che la frequenza dei proverbi è meno elevata di quella delle espressioni idiomatiche o di altri tipi di sequenze fisse, sebbene lamentiamo una certa carenza di studi statistici che possano dare conto del comportamento anche quantitativo dei proverbi. Un metro di paragone che offre almeno un'idea della proporzione tra quantità di parole nel discorso libero e numero di proverbi effettivamente utilizzati nella comunicazione è fornito da Norrick (1985: 6), che, analizzando un corpus di conversazioni in inglese (il corpus in questione è *A corpus of English Conversation*) di 43.165 righe, ossia 891 pagine (purtroppo non fornisce la quantità dei token presenti), ha trovato un solo proverbio. Per questa ragione, se si volessero studiare i proverbi nella conversazione, servirebbe una mole considerevole di dati. Dal canto nostro, nella fase di scelta del corpus, abbiamo da subito ristretto la scelta a corpora di grandi dimensioni, dato che questo sembra essere un prerequisito allo studio dei proverbi in corpora linguistici. Ciononostante, non ci siamo concentrati solo sulla quantità di dati presenti nei corpora, ma anche sulla qualità dei corpora stessi.

Oltre a corpora di grandi dimensioni e qualitativamente validi, in quanto rappresentativi di un ampio spettro di fenomeni linguistici più o meno frequenti, ci siamo concentrati sulla possibilità di lavorare su almeno due corpora comparabili, uno di lingua italiana e l'altro di lingua francese. Come si vedrà, anche da questo punto di vista i corpora WaCky sono estremamente interessanti. In particolare, ci concentreremo sul rapporto tra itWaC e frWaC. Come si diceva, la modalità di costruzione dei due corpora sono le medesime. Il corpus itWaC è stato ottenuto utilizzando come *seeds* 1.000 coppie di parole scelte casualmente tra le parole di frequenza del corpus de *la Repubblica* e la lista di vocabolario di base. Il dominio di ricerca automatica è stato limitato ai siti col suffisso .it. Parallelamente, il corpus frWaC è stato ottenuto utilizzando come *seeds* 1.000 coppie di parole scelte casualmente tra le parole di frequenza media del corpus di *Le Monde Diplomatique* e la lista di vocabolario di base. Anche in questo caso, si è deciso di limitare la ricerca ai soli siti con suffisso .fr. Capiamo da subito che i criteri di scelta dei *seeds*, che potremmo considerare le fondamenta dei due corpora, sono assolutamente comparabili.

Si è già parlato di come sono stati puliti i corpora (cfr. § 2.3) per evitare duplicati, totali o parziali, nonché gli altri tipi di rumore, come possono essere i testi scritti in lingue diverse da quella del relativo corpus, forestierismi, errori di battitura, ecc. Queste operazioni

sono dunque state effettuate seguendo gli stessi principi per tutti i corpora WaCky. Pertanto, anche la fase di pulitura dei corpora itWaC e frWaC è, a tutti gli effetti, comparabile.

L'etichettatura morfo-sintattica è stata realizzata impiegando il software TreeTagger³⁸ per entrambi i corpora presi in esame. Per quanto concerne la lemmatizzazione di itWaC, tale operazione è stata effettuata mediante il software Morph-it³⁹, mentre per frWaC è stato sempre utilizzato TreeTagger. Nonostante questa piccola differenza, possiamo considerare i due corpora comparabili anche dal punto di vista dell'annotazione.

Vediamo ora qualche statistica generale relativa ai due corpora. Il corpus itWaC contiene quasi due miliardi di *token*, per la precisione 1.909.698.363; il corpus frWaC è leggermente più ridotto, essendo composto da un miliardo e seicento milioni circa di *token*, per la precisione 1.613.206.614. Il numero di frasi è di conseguenza maggiore in itWaC (68.147.599) che in frWaC (54.875.342). Possiamo asserire che, in media, i documenti contenuti in itWaC hanno una lunghezza maggiore di quelli di frWaC, dato che, sebbene il numero di *token* sia maggiore in itWaC, la quantità di documenti di frWaC supera quella di itWaC. In frWaC vi sono più di 2 milioni e 250 mila documenti, mentre in itWaC ve ne sono quasi un milione e 900 mila. Ciò significa che i confini quantitativi tra i corpora presi in esame sono alquanto labili, considerata l'impossibilità empirica di creare due corpora comparabili in due lingue diverse aventi esattamente lo stesso numero di *token*, di frasi o di documenti se non tramite interventi diretti da parte dell'autore nei confronti del corpus che non avrebbero altro effetto che alterarne la validità.

Date le caratteristiche simili a livello di costruzione, questi due corpora possono essere considerati comparabili: certamente, differiscono dai classici corpora comparabili per il fatto che non fanno riferimento a uno o più ambiti terminologici, bensì alla lingua generale, ma essendo così strettamente comparabili, potrebbero dare conto di fenomeni linguistici della stessa natura senza che essi siano l'uno la traduzione dell'altro. In altre parole, l'analisi dei proverbi effettuata prima in un corpus e poi nell'altro seguendo le stesse regole e i medesimi procedimenti può fornire risultati molto importanti, in quanto assolutamente comparabili ma al contempo oggettivi e autentici, poiché non falsati dalla mano del ricercatore o dalla presenza di processi traduttivi. Da un'altra prospettiva, questi corpora danno la possibilità non solo di comparare gli oggetti ricercati ma anche di condurre ricerche *corpus-driven* (cfr. § 2.7),

³⁸ Cfr. <http://www.ims.uni-stuttgart.de/projekte/complex/TreeTagger/>.

³⁹ <http://sslmit.unibo.it/morphit>.

e quindi *bottom-up*, partendo da dati reali per giungere alla formulazione di ricorrenze e tendenze generali, fino all'individuazione del comportamento semantico degli oggetti studiati.

2.5. Strumenti di consultazione dei corpora WaCky

Per consultare i corpora WaCky e sfruttarne appieno le potenzialità, vi sono diversi strumenti, anche gratuiti, come interfacce e piattaforme. Nell'ambito del presente studio, si è deciso di impiegare uno strumento di analisi di corpus e di realizzazione di concordanze basate sul web chiamato *NoSketch Engine*⁴⁰. Questa risorsa è ospitata dal *Natural Language Server*⁴¹ del Dipartimento di Tecnologie della Conoscenza (*Dept. Of Knowledge Technologies*) presso l'Istituto Jožef Stefan (Lubiana, Slovenia). Nell'interfaccia di concordanza di questo sito si trovano i corpora WaCky ad oggi esistenti, selezionabili facilmente da un menù a tendina. Le ricerche nei corpora si effettuano utilizzando il cosiddetto *Corpus Query Language* (CQL), un linguaggio di ricerca che prende spunto dal linguaggio sviluppato presso il *Corpora and Lexicons group*, agli inizi degli anni '90 del secolo scorso, modificato successivamente sotto diversi aspetti. L'ultimo aggiornamento della versione del CQL da noi utilizzata risale all'aprile del 2015.⁴²

In riferimento all'annotazione dei corpora, sebbene sia stato utilizzato il software TreeTagger per etichettare morfo-sintatticamente i due corpora, è stato necessario reperire la nuova lista delle etichette (*tagset*) utilizzate, che differiscono da quelle impiegate in TreeTagger non dal punto di vista funzionale, ma semplicemente dal punto di vista formale (le etichette utilizzate in TreeTagger sono state rinominate). La lista delle etichette utilizzate in itWaC è reperibile al § 3.5.17 di questo link: <http://nl.ijs.si/spook/msd/html-sl/msd-it.html>. Da notare, infine, che i *tag* utilizzati per itWaC non corrispondono formalmente a quelli impiegati per frWaC, i quali sono elencati nella tabella del § 3.4.16 all'indirizzo <http://nl.ijs.si/spook/msd/html-sl/msd-fr.html>.

2.6. Linguistica dei corpora e paremiologia

L'approccio proposto nell'ambito della presente Tesi di Dottorato è volto ad applicare i principi della linguistica dei corpora alla paremiologia. Sarà, pertanto, necessario reperire i

⁴⁰ Per maggiori informazioni o per utilizzare questa risorsa, si faccia riferimento al seguente link: <http://nl.ijs.si/noske/index-en.html>.

⁴¹ Si veda: <http://nl.ijs.si/>.

⁴² Per informazioni più dettagliate circa questo linguaggio, si segnala il link <https://www.sketchengine.co.uk/xdocumentation/wiki/SkE/CorpusQuerying>.

proverbi nei corpora appena descritti per tentare di individuare tendenze generali relative al proverbio e al suo uso.

Il punto di partenza deve essere chiaro da subito e sgombro da eventuali equivoci. Quanto si auspica di ottenere grazie a questo lavoro è anzitutto una serie di osservazioni oggettive e scientifiche riguardanti il proverbio basate esclusivamente su dati reali, ossia dati prodotti spontaneamente dai parlanti e non manipolati dal ricercatore. Pertanto, il proverbio non sarà più visto, d'ora in poi, come un elemento linguistico a sé stante, ovvero come frase, bensì come oggetto esclusivamente linguistico enunciato dai membri di una comunità in contesti spontanei e non in ambiti legati al folklore oppure per mere finalità metaparemiologiche. Nel presente studio, pertanto, il proverbio verrà considerato un enunciato che viene impiegato per veicolare messaggi in contesti comunicativi.

La prima domanda che è lecito porsi è la seguente: è possibile reperire proverbi in testi? Secondo il grande paremiologo ungherese Paczolay Gyula, autore tra gli altri del recentissimo dizionario di proverbi ungheresi tradotti in numerose lingue (Paczolai, 2015), non è possibile trovare proverbi in testi⁴³. Almeno, non è questa la base di partenza. Il proverbio esiste prima nelle opere folkloristiche e lessicografiche e poi, di riflesso, nella comunicazione. Pertanto, un testo non potrà discernere tra ciò che è un proverbio e ciò che non lo è; dunque, non lo si può interrogare per ottenere come risultati proverbi o informazioni circa l'uso dei proverbi. La nostra risposta, invece, è diametralmente opposta: è possibile, anzi, doveroso, reperire proverbi in testi. Questo per molte ragioni diverse: anzitutto, perché la paremiologia deve entrare a far parte a tutti gli effetti delle principali discipline linguistiche e deve iniziare ad appropriarsi degli strumenti di analisi già da tempo impiegati in linguistica, così come in una delle sue branche come la fraseologia. In secondo luogo, perché è incluso nell'essenza stessa del proverbio, che è frase, e quindi un potenziale enunciato o testo, di essere impiegato in situazioni comunicative reali.

La seconda domanda che sorge è la seguente: posto che è possibile ritrovare proverbi in testi, come si procede al loro reperimento? Ebbene, i procedimenti possono essere divisi in due categorie, adattate dalla ricerca relativa alle espressioni idiomatiche, ambito sul quale sono stati condotti studi più numerosi rispetto a quelli condotti sui proverbi.

⁴³ Comunicazione personale nell'ambito del convegno ICP15 tenutosi a Tavira (Portogallo) dal 1° all'8 novembre 2015.

La prima modalità potrebbe consistere nel cercare all'interno di corpora linguistici proverbi preselezionati sulla base di liste o dizionari. Sicuramente questo procedimento è molto rapido, tuttavia non lo si può considerare altrettanto valido. La primissima critica che si può avanzare nei confronti di tale modalità è la non certezza a priori di ritrovare nei corpora gli elementi individuati, in quanto essi potrebbero non essere più usati nella comunicazione. Come osserva Moon (1998: 47), le prime ricerche sulle espressioni idiomatiche si basavano effettivamente su liste di elementi conosciuti, estratti da testi o dizionari senza criteri ben definiti; il rischio che si presentava spesso era quello di imbattersi in opere relative a espressioni idiomatiche che mettessero insieme e senza distinzioni espressioni idiomatiche non effettivamente utilizzate perché rare o desuete con espressioni idiomatiche comuni o frequenti. Molto spesso, infatti, mancavano importanti annotazioni sulla frequenza d'uso dei proverbi.

La seconda modalità, invece, consisterebbe nel reperire direttamente nel corpus i proverbi. In questo secondo caso il corpus diventa, in una prima fase, fonte di proverbi e, in una seconda fase, strumento attraverso il quale analizzare i proverbi in esso ritrovati, seppur con possibili modifiche o ampliamenti del materiale da analizzare. Rispetto al primo procedimento, questo è più scientifico, dato che non si influisce sul materiale da studiare nemmeno nella fase di reperimento dello stesso, evitando di concentrarsi, anche in maniera involontaria, solo su alcuni elementi per trascurarne altri. Tuttavia, emerge da subito la difficoltà di un simile procedimento. Com'è possibile, infatti, reperire in un corpus proverbi di cui si ignora tutto tranne il fatto che devono essere proverbi? Ad oggi, per quanto ne sappiamo, non esistono sistemi di annotazione di corpora che annotino e identifichino i proverbi.

2.7. Verso un approccio *corpus-driven* all'analisi paremiologica

Prima di passare alla nostra proposta relativa al reperimento e all'analisi dei proverbi all'interno di corpora, è opportuno aprire una parentesi sui due principali approcci utilizzati, a tal proposito, in linguistica dei corpora: l'approccio *corpus-based* e l'approccio *corpus-driven*.

Il primo dei due approcci vede la formulazione di teorie e descrizioni linguistiche al di fuori dei corpora linguistici; come erano soliti fare i grammatici nel passato, l'approccio *corpus-based* consiste nella formulazione di teorie linguistiche in base alla propria esperienza acquisita in ambito linguistico. La tappa successiva consiste nel cercare conferme alle ipotesi formulate grazie all'intuizione di parlanti nativi o con grande competenza. I corpora, in questo

ambito, possono sostituire il confronto con parlanti nativi, fornendo prove ancora più certe, veritiere e quantificabili. Riassumendo, lo scopo dell'approccio *corpus-based* è di verificare ipotesi formulate a priori. Trattando tale approccio, Tognini-Bonelli (2001: 15-18) presenta l'esempio seguente riguardante la preposizione *any* in inglese. Le grammatiche solitamente indicano che questa preposizione viene impiegata in frasi negative, interrogative, dopo *if/whether* (quindi in costrutti condizionali) e in espressioni di dubbio. Ebbene, la novità che ha apportato la linguistica dei corpora con il suo approccio *corpus-based* è stata quella di rendere possibile l'interrogazione dei corpora per verificare e quantificare teorie linguistiche esistenti, come quella riportata nell'esempio, utilizzando i dati dei corpora. Tuttavia, tale approccio comporta alcuni limiti non del tutto trascurabili. Sempre considerando l'esempio di Tognini-Bonelli (*ibid.*), i corpora utilizzati per verificare le ipotesi previamente espresse mostrano quanto queste possano essere valide, ma al contempo indicano che vi è tutta una serie di casi che esula da qualsiasi teoria esistente. Nell'esempio di *any* mostrato da Tognini-Bonelli (*ibid.*), emerge che, secondo il corpus consultato, quasi nella metà dei casi *any* risulta impiegato in frasi affermative, cosa che non trova conferma o spiegazione nelle grammatiche tradizionali.

Si noterà che i corpora, utilizzati con questo procedimento che potremmo definire *top-down*, ovvero per verificare teorie elaborate in precedenza con dati concreti, non possono esprimere tutto il loro potenziale.

Entra in gioco, a questo punto, il procedimento *corpus-driven*, grazie al quale il ricercatore non parte da teorie preconfezionate, bensì cerca di interrogare i corpora consultati affinché siano questi a evidenziare una teoria. Se il precedente approccio poteva essere considerato *top-down*, ovvero da un'ipotesi formulata a priori si cercavano evidenze nei dati a disposizione del ricercatore, l'approccio *corpus-driven* è *bottom-up*: partendo dai dati forniti dal corpus si formulano ipotesi sulla base di pattern ricorrenti che emergono nel corso dell'analisi. Tali ipotesi portano a una generalizzazione in termini di regole e uso per poi permettere la formulazione di una teoria (Tognini-Bonelli, 2001: 17). Riprendendo l'esempio di *any*, se l'approccio *corpus-based* aveva confermato solo in parte la teoria precedentemente espressa, l'approccio *corpus-driven* porta il ricercatore a un nuovo paradigma che sia in grado di motivare quei casi non illustrati dalla teoria ma ricorrenti nel corpus.

Notiamo che il procedimento *corpus-driven*, per poter giungere a risultati interessanti, passa dall'individuazione di pattern ricorrenti sulla base dei quali proporre una

generalizzazione dei fenomeni osservati e, dunque, formulare una teoria. Si rende necessario, a questo punto, capire meglio che cosa si intenda per pattern. Come ricorda Tognini-Bonelli (2001: 90), la nozione di pattern è al confine tra il lessico e la grammatica ed è al centro dell'approccio *corpus-driven*. Il pattern di una parola altro non è che l'insieme delle parole e delle strutture che accompagnano regolarmente una parola. A questo punto, non ci rimane che capire come si possa individuare un pattern ma anche a che cosa permettano di giungere i pattern individuati. Definendo la nozione di pattern si è impiegato, non a caso, un avverbio di grande importanza in questo ambito: "regolarmente". Affinché sia possibile formulare generalizzazioni o teorie partendo dai dati raccolti nei corpora, il ricercatore o il discente non si può limitare a reperire un solo caso in cui appare un certo pattern, bensì deve trovare una serie di casi in cui si ripete un certo pattern affinché questo possa essere rilevante, vale a dire non dettato dalla casualità, ma supportato da prove (la frequenza). L'approccio *corpus-driven*, pertanto, si basa sulla ricerca e sull'analisi di eventi ripetuti. Per questa ragione, in linguistica dei corpora si cerca una o più parole in un corpus, utilizzando uno strumento chiamato *concordanza*. Grazie alla concordanza è anzitutto possibile estrarre le parole desiderate, che, una volta reperite, diventano il *nodo* della concordanza, vale a dire la parola chiave della concordanza mostrata in tutti i diversi casi in cui viene impiegata nei testi del corpus, con i rispettivi cotesti. Grazie alla concordanza è possibile ottenere in tempi rapidissimi informazioni molto rilevanti, come l'occorrenza del nodo, ovvero quante volte si ripete il nodo nell'intero corpus. Solitamente, il cotesto viene visualizzato nella forma del cosiddetto KWIC (*keyword-in-context*), con il quale si mostrano le parole precedenti e successive al nodo, scegliendo la quantità di tali parole oppure la modalità con cui queste devono essere visualizzate (ad esempio, decidendo di visualizzare la frase intera) (Chiari, 2007: 72-73). Il nodo viene, così, evidenziato nella parte centrale della schermata e ogni esempio viene isolato in una riga per non confonderlo con gli esempi successivi o precedenti. È possibile mantenere l'ordine con cui la parola chiave si presenta nel testo oppure scegliere di ordinare gli esempi a seconda di una o più parole a destra o a sinistra del nodo stesso. Il tutto a completa discrezione dell'utente del corpus. Una volta che questi abbia ordinato a suo piacere gli esempi ottenuti in maniera da far emergere più chiaramente possibile i pattern da lui cercati, l'utente del corpus può procedere all'analisi degli esempi in cerca di ricorrenze.

La lettura delle concordanze è un'operazione che necessita di una competenza molto elevata da parte dell'utente del corpus. Infatti, né il corpus né i software utilizzati per la

gestione dei corpora possono arrivare da soli a formulare teorie. Tutto quanto viene impiegato in linguistica dei corpora è solamente uno strumento che deve servire al ricercatore o al discente per giungere a determinate conclusioni sulla base degli esempi verificati nei corpora in loro possesso e su dati riguardanti la qualità e la quantità dei pattern individuati.

Ogni riga presente in una concordanza è un esempio isolato di lingua in uso che potrebbe anche non ripetersi mai più nella forma esatta in cui compare in quella determinata riga. In termini saussuriani, si potrebbe pertanto sostenere che ogni riga sia un esempio, una realizzazione, della *parole*. Il ricercatore, una volta ottenuta una concordanza, procede alla lettura verticale della parola chiave in contesto, alla ricerca di caratteristiche formali condivise da raggruppare e analizzare. Tale tipo di lettura verticale è di fondamentale importanza per l'individuazione di pattern ricorrenti, i quali sono altresì pattern della *langue*. Ecco che la linguistica dei corpora ha portato a una rivoluzione, ossia all'osservabilità della *langue* attraverso l'analisi di insiemi sempre più nutriti di esempi di *parole*, un fatto impensabile prima dell'avvento dei corpora elettronici (Tognini-Bonelli, 2001: 98-98).

2.8. Esperienze di applicazione della linguistica dei corpora alla fraseologia

Terminiamo questa panoramica sulla linguistica dei corpora e sulla paremiologia proponendo alcuni studi effettuati nell'ambito della metafora. Sebbene si tratti di un'altra area di ricerca, abbiamo notato che tali studi possono essere grande utilità ai fini del presente lavoro, in quanto evidenziano problematiche valide anche per la paremiologia e propongono soluzioni che potrebbero essere ugualmente applicate all'analisi del proverbio.

Il primo di questi ambiti è quello della metafora. Senza entrare nel merito della metafora e degli studi sulla metafora, cerchiamo di capire meglio quali siano gli aspetti in comune tra la metafora e i proverbi. Partiamo da una ricerca realizzata da Deignan (2009) sui pattern metaforici nei corpora. Precisiamo che la metafora considerata in questo studio non è quella poetica, quindi creativa, bensì la metafora come intesa e descritta nella *Conceptual Metaphor Theory* di Lakoff e Johnson (1980), vale a dire un tipo di metafora morta e convenzionale che si è oramai acclimatata nella lingua e nel pensiero di una data comunità. Questo tipo di metafora, detta anche *catacresi* (Cfr. Prandi, 2006: 370-375), è una metafora che passa quasi inosservata quando viene utilizzata tanto è ormai entrata nell'uso comune. Essa, diversamente dalla metafora creativa, non è conflittuale, ma anzi coerente; pertanto, è entrata a pieno titolo nel patrimonio condiviso di concetti di una data comunità. Esempi di

catacresi sono il *denaro liquido*, oppure *l'ala di un edificio* o ancora *l'amore che scoppia tra due individui*.

Una delle problematiche che emergono in Deignan (2009) riguarda il reperimento di tali metafore in contesti d'uso. L'utilizzo di corpora, per la natura stessa degli obiettivi attesi, appare scontato, ma non sembra altrettanto lampante la metodologia da seguire per individuare tali metafore. Deignan (2009: 15) rimarca non solo la fondamentale capacità della linguistica dei corpora di permettere uno studio accurato tanto sulle metafore quanto sui contesti in cui esse vengono impiegate, ma evidenzia anche che il migliore approccio da seguire è quello *corpus-driven*. Tale procedimento, non a caso, è l'unico di quelli solitamente seguiti in linguistica dei corpora che è in grado di prendere in considerazione la totalità dei dati disponibili in un corpus rispondenti a certi criteri stabiliti a priori, senza che il reperimento dei dati sia condizionato da ipotesi prestabili o da esempi cercati ad hoc, che rischierebbero di rappresentare una sezione ristretta dell'intera classe di fenomeni da analizzare. Deignan (2009: 14) porta un esempio a sostegno di questa sua posizione. La *Conceptual Metaphor Theory* propone pattern che vengono spesso validati dall'analisi dei contesti, ma non sempre contempla altri pattern che invece emergono grazie a un'accurata analisi dei dati. Prendiamo come esempio *l'argomentazione* che, secondo la teoria appena citata, è *guerra* (nei termini della stessa teoria: ARGUMENT IS WAR). Per questa ragione si possono *attaccare e difendere* tesi, teorie, argomentazioni, etc. In altri termini, se prendiamo, da un lato, il dominio dell'argomentazione e, dall'altro lato, il dominio della guerra, notiamo che instaurano rapporti lessicali paralleli. Tuttavia, Deignan (2009: 14-15) ha analizzato le occorrenze del verbo *attack* e successivamente del sostantivo *strategy*. Ciò che ha potuto notare è che sia per il verbo in esame sia per il sostantivo non vi è solamente un campo di uso possibile, ossia quello della guerra, bensì una scala di impieghi di volta in volta più o meno vicini al campo della guerra. Si è potuti giungere a tali osservazioni solo grazie alla scelta dell'approccio *corpus-driven*, con il quale si è davvero interrogato il corpus non tanto per verificare un'ipotesi preesistente, quanto piuttosto per individuare i principali pattern semantici dei due elementi cercati. Tale procedimento, dunque, può generare nuove ipotesi e teorie sulla base dell'analisi dei dati, cosa che non sarebbe altrettanto possibile utilizzando l'approccio *corpus-based*, con il quale si procede a una mera verifica di teorie esistenti, tralasciando in partenza i casi che non possono essere spiegati della teoria stessa.

Tuttavia, resta da sciogliere un nodo non certo di poca importanza: com'è possibile reperire metafore nei corpora? Deignan (2009), nel § 2.4.2 del suo articolo, cerca di dare una risposta a tale quesito, mostrando anzitutto quanto già è stato fatto. Il primo metodo che potrebbe venire in mente per reperire tutte le metafore contenute in un corpus consiste nel consultare manualmente il corpus stesso alla ricerca di metafore. Si tratterebbe probabilmente del metodo più accurato per eseguire il compito, ma la realtà dei fatti, ovvero la mancanza di tempo e risorse per fare ciò, restringe l'area di applicazione di questo procedimento ai soli corpora di piccole dimensioni. Nel caso di corpora di grandi dimensioni si potrebbe ricorrere alla selezione delle metafore da cercare nel corpus, ma così facendo si rischierebbe di tagliare fuori dall'analisi tutta una serie di dati che potrebbero al contrario essere di grande interesse per la ricerca. Un'alternativa a questi due approcci, nonché una loro possibile fusione, potrebbe consistere nell'analizzare campioni di corpus al fine di individuarvi parole o espressioni ricorrenti in costrutti metaforici da utilizzare in una fase successiva per cercare ulteriori concordanze all'interno del corpus. Gli svantaggi di questo approccio stanno nell'impossibilità di reperire tutte le metafore presenti in un corpus, nonché nella difficoltà di garantire un'alta affidabilità del sistema nel caso in cui il corpus sia composto da testi troppo disparati.

Per riuscire a portare a termine l'impresa, cercando di rispettare la logica dell'approccio *corpus-driven*, si può cercare una metafora non cercando direttamente la stessa metafora, molto spesso imprevedibile, bensì un qualcosa che con alta frequenza la accompagni. In tal senso, Stefanowitsch (2006) ritiene che non sia errato trovare una metafora indirettamente, ovvero individuando previamente i cosiddetti marcatori della metafora, che altro non sono che espressioni, come *una sorta di (kind of)*, *per così dire (so to speak)*, che segnalano la presenza di una metafora. Anche questo procedimento, tuttavia, ha determinati limiti, in quanto non prende in considerazione tutti quei casi in cui la metafora appare da sola, senza essere segnalata da alcun marcatore.

Deignan (2009: 17-19) segnala, poi, una serie di esperienze già messe a punto relative all'identificazione automatica delle metafore in un corpus. Vi è un filone di progetti sviluppati per individuare quelle parole tipiche di certi campi semantici ma presenti in testi appartenenti ad altre aree semantiche in cui non dovrebbero essere presenti. Ad esempio, il verbo *pour* (versare) colloca con il sostantivo *liquid* (liquido) nel dominio della chimica di laboratorio, pertanto quando lo si trova collocato con altri sostantivi che non appartengono a questo

dominio, ad esempio un sostantivo finanziario come *assets* (risorse, attivi finanziari), verrà utilizzato in maniera metaforica. In questo ambito ci si avvale anche di concetti quali la *distanza semantica*, che indica quanto distano semanticamente due collocati e quindi la probabilità che le parole in questione vengano usate in maniera metaforica. Tali procedimenti, tuttavia, fanno maggiormente riferimento all'uso di parole singole e quindi alle collocazioni e sono di difficile applicazione nei confronti di espressioni idiomatiche o proverbi, in quanto, le espressioni idiomatiche o i proverbi sono quasi sempre coerenti nel loro significato compositivo.

2.9. Come individuare automaticamente proverbi in corpora?

Come si diceva in precedenza, il lavoro in oggetto è volto a proporre uno studio della semantica del proverbio italiano e francese, intendendo con ciò un'analisi che possa rendere conto degli usi e delle modifiche di cui sono oggetto tanto i proverbi italiani quanto quelli francesi, basandosi unicamente sullo studio di proverbi reperiti nei corpora poc' anzi descritti, ossia itWaC e frWaC. L'ottica adottata si rifà a quella di un professionista delle lingue, tipicamente un interprete o un traduttore, che si trova di fronte alla traduzione interlinguistica di proverbi e che deve operare la propria scelta traduttiva tra i soli proverbi effettivamente in uso in L2, con un significato paremiologico corrispondente e con parametri di frequenza e notorietà simili. È chiaro, dunque, che si predilige in questa ricerca una prospettiva comparata: in altri termini, si effettuano studi dello stesso tipo prima nel corpus di L1 (itWaC) e poi nel corpus di L2 (frWaC), aventi tra loro rapporti di corrispondenza quanto ai criteri di costruzione dei corpora, ma senza essere l'uno la traduzione dell'altro. Visto sotto un'altra ottica, non si possono analizzare corpora paralleli, ossia corpora che contengono traduzioni dei testi presenti in altri corpora, per non falsare quei criteri essenziali ai fini del presente studio, che prevedono che l'analisi dei proverbi sia effettuata sulla base di corpora di almeno due lingue distinte, seguendo una metodologia *corpus-driven*. Se un corpus fosse la traduzione dell'altro, tali principi di corrispondenza di costruzione dei corpora verrebbero a meno, il che minerebbe le basi dell'intera ricerca. Per tutte queste ragioni, dunque, si è ritenuto necessario utilizzare unicamente corpora comparabili e di grandi dimensioni, per garantire il reperimento del maggior numero possibile di proverbi.

Per fornire un'analisi quanto più ampia, generale ed esaustiva possibile, si cercherà altresì di concentrare la propria attenzione sui proverbi di tutti i tipi, senza fare distinzione alcuna tra i proverbi che la tradizione considera metaforici o letterali (distinzione che non

accettiamo per quando abbiamo esposto in § 1.3.1 e in nota 18), ma nemmeno distinzioni di carattere semantico (considerando uno qualunque dei livelli di significato del proverbio, da quello riguardante il significato compositivo a quello relativo al suo significato paremiologico). Si delinea, pertanto, lo scopo di questa fase iniziale di ricerca dei proverbi che consiste nel reperimento del maggior numero di proverbi effettivamente utilizzati da parlanti nativi italiani e francesi. Per le qualità dei corpora a nostra disposizione, possiamo aggiungere alcune caratteristiche che a priori dovrebbero presentare i proverbi da reperire: essi dovrebbero essere proverbi effettivamente utilizzati nei primi anni 2000 (ovvero nel periodo di costruzione dei corpora anche se, come sottolineeremo in § 3.5, c'è un piccolo slittamento della datazione media dei testi appartenenti ai due corpora) dagli utenti della Rete, comprendendo non solo gli autori di articoli di giornali e riviste, ma anche i partecipanti a discussioni o autori di tipologie testuali riconducibili a quelle generalmente presenti sui blog.

Prima di illustrare le tappe raggiunte nella ricerca dei proverbi nei nostri due corpora, occorre evidenziare che, per individuare la metodologia di reperimento dei proverbi in contesto, si è utilizzato unicamente il corpus che abbiamo battezzato di L1, ovvero itWaC. Per ragioni di tempo e praticità, abbiamo considerato superfluo testare la stessa metodologia in entrambi i corpora, considerata la loro comparabilità. È molto probabile, infatti, che a un fallimento in termini di metodologia sperimentata sul corpus di L1 (itWaC) sarebbe corrisposto un simile insuccesso nel corpus L2 (frWaC). In realtà, crediamo che i fallimenti da noi riscontrati, così come i successi da noi verificati, possano essere validi anche per altri corpora generali o di riferimento, in quanto essi sono da imputare alle caratteristiche intrinseche dei proverbi, dell'intero patrimonio paremiologico e di come essi si presentano nell'uso.

Procediamo, dunque, all'enunciazione dei principi e dei tentativi da noi seguiti per individuare una metodologia applicabile a entrambi i corpora. La nostra esperienza di applicazione dei principi della linguistica dei corpora ai proverbi parte da una serie di tentativi ideati e applicati al corpus itWaC, aventi la finalità di reperire automaticamente proverbi di qualsiasi tipo nello stesso corpus.

Come si è già ricordato, è una vera e propria sfida riuscire a identificare proverbi in un corpus così vasto. Probabilmente, è addirittura più complicato che recuperare espressioni idiomatiche o altre forme di fissità linguistica, e questo per diverse ragioni. Anzitutto perché, in base a quanto si è dimostrato nel capitolo riguardante la linguistica del proverbio, il

proverbio è una frase mentre l'espressione idiomatica è un costituente. Tale differenza fondamentale porta ad avere diverse caratteristiche enunciative: i proverbi, rispetto alle espressioni idiomatiche, presentano molte più varianti che frequentemente non sono nemmeno prese in considerazione dai dizionari di proverbi; i proverbi godono di uno status citazionale che fa sì che il parlante li enunci come se fossero una qualsiasi citazione riportata in un testo. Sotto un'altra ottica, si potrebbe sostenere che i proverbi godano di uno status di fissità diverso rispetto alle espressioni idiomatiche.⁴⁴ Qualunque sia il punto di vista, tale caratteristica è oggettiva ma soprattutto ha ricadute di grande rilevanza per il reperimento dei proverbi in un corpus, non solo automatico, ma anche manuale.

In effetti, è da tener presente che, sebbene un'espressione idiomatica sia facilmente reperibile in un corpus inserendo semplicemente nella stringa di ricerca le parole che la compongono⁴⁵, la stessa modalità di ricerca non è detto che porti a risultati sempre felici quando si lavora con i proverbi. È sufficiente che il proverbio cercato sia intervallato da parole a esso estranee oppure che la variante cercata non sia quella attestata nel corpus perché il proverbio stesso non venga reperito. Nonostante tale consapevolezza, sarebbe comunque un enorme passo avanti capire le ragioni per le quali di volta in volta non emergono i risultati sperati. Ebbene, il metodo *corpus-based*, con cui si cercano proverbi noti in un corpus, non fornisce elementi sufficienti per capire la causa di un fallimento nel caso del non reperimento dei proverbi. Per tutte queste ragioni si può supporre che un approccio *corpus-based* alla ricerca del proverbio, oltre a produrre risultati più limitati rispetto a una ricerca simile *corpus-driven* poiché parte da un insieme dato di elementi, si possa rivelare controproducente dato che potrebbe nascondere risultati senza fornire al ricercatore la possibilità di capirne la ragione.

⁴⁴ Alcuni autori che parlano di maggiore o minore fissità. Ora, non è nostra intenzione entrare nel merito di questa disquisizione, non essendo questa la sede più appropriata. Ciononostante, è indubbio che il tipo di fissità che contraddistingue proverbi ed espressioni idiomatiche (in senso lato) sia molto diverso, specialmente per quanto concerne il discorso di *déproverbialisation* messo in luce da Schapira (2000) che riguarda principalmente i proverbi, ma soprattutto stando al tema delle varianti di proverbi (a volte, nemmeno i dizionari danno conto di tutte le varianti esistenti di un proverbio) nonché allo status citazionale dei proverbi che fa sì che l'enunciazione del proverbio sia spesso intervallata da altro che non fa parte del proverbio stesso. Tutto questo si ricollega al concetto più generale e fondamentale espresso all'inizio di questa tesi per il quale il proverbio è frase, mentre l'espressione idiomatica è costituente di frase.

⁴⁵ È doveroso, tuttavia, ricordare che Moon (1998: 47), che si è a lungo occupata di fraseologia all'interno della *corpus linguistics*, riferendosi alle espressioni idiomatiche evidenzia come il concetto stesso di fissità debba essere rimesso in discussione sulla base delle evidenze tratte dai corpora, i quali "show up clearly the fallacy of the notion of fixedness of form". Ciò significa che l'opinione comune secondo la quale le espressioni idiomatiche siano altamente fisse non corrisponde sempre alla realtà.

Si ritorna, dunque, al quesito posto in precedenza: com'è possibile trovare proverbi all'interno di corpora di così grandi dimensioni? L'approccio migliore consisterebbe nell'averne un corpus in cui siano previste etichette specifiche per i proverbi, ma per ora questa strada non è ancora percorribile.

Scartata in partenza tale metodologia, abbiamo tentato di individuare i proverbi seguendo una lista già redatta, sebbene fossimo consapevoli dei limiti legati a tale approccio, sicuramente più affine alla metodologia *corpus-based* che a quella *corpus-driven*. I proverbi appartenenti alla lista sarebbero stati cercati nel corpus uno a uno, ottenendo così le concordanze di ogni singolo proverbio. Ciononostante, tale metodologia avrebbe messo in evidenza almeno due ostacoli, entrambi scaturiti dalle modalità di redazione della lista di proverbi da cui partire. Anzitutto, tale lista avrebbe dovuto tenere conto delle varietà di proverbi più comuni e maggiormente presenti in testi e discorsi. Come si diceva prima, trovare le concordanze di un proverbio significa *in primis* trovare l'esatta combinazione di parole che dovranno essere presenti nel corpus. Se così non fosse, si correrebbe il rischio di non giungere ad alcun risultato. L'altro problema di grande rilevanza da risolvere prima di comporre la lista consisteva nel concentrarsi unicamente su quei proverbi effettivamente utilizzati ai nostri giorni. Le liste di proverbi che si possono reperire in molti luoghi diversi (non solo in opere, compendi e dizionari di proverbi, ma anche in Rete) non sempre tengono conto della realtà, vale a dire dei proverbi che vengono effettivamente impiegati in testi e discorsi. È quanto osserva anche Moon (1998: 45) che rileva che "most such sources - general dictionaries, or specialist dictionaries of idioms - record and perpetuate items not necessarily found in current English". Questa ipotesi, a priori potenzialmente valida anche per l'italiano, è stata confermata da una nostra esperienza concreta (cfr. § 2.9.1 e 2.9.2).

Inoltre, ci siamo dovuti preoccupare anzitutto di eliminare qualsiasi influsso della nostra soggettività nei confronti della lista di proverbi da studiare. In effetti, volendo produrre uno studio scientifico, il più possibile valido per altri proverbi non direttamente studiati in questa sede, abbiamo ritenuto indispensabile arrivare a una lista di proverbi non basata su scelte nostre o di un campione di parlanti. Certo, questa metodologia sarebbe stata maggiormente affine a un approccio *corpus-based*, ma ci avrebbe senz'altro permesso di ottenere elementi eterogenei e relativi a tutte le diverse tipologie di proverbi possibili.

Per queste ragioni, abbiamo da subito deciso di scegliere casualmente i proverbi. Dato che il nostro obiettivo consisteva anzitutto nel prendere in considerazione la quantità

maggior possibile di tipologie di proverbi italiani, abbiamo deciso di utilizzare alcune tra le principali collezioni di proverbi italiani, restringendo la nostra ricerca alle soli fonti cartacee in quanto ritenute più affidabili e più complete delle risorse italiane online. In particolare, la nostra scelta si è limitata ai seguenti tre dizionari di proverbi: *Il grande dizionario dei proverbi italiani* (Guazzotti e Oddera, 2006), il *Dizionario dei proverbi italiani* (Lapucci, 2006) e il *Dizionario dei proverbi* (Boggione e Massobrio, 2004). Cerchiamo di presentare le caratteristiche di ognuno di questi dizionari.

2.9.1. Tre dizionari di proverbi italiani a confronto

I tre dizionari sopramenzionati costituiscono alcune delle opere di riferimento per la paremiologia italiana, in quanto riassumono adeguatamente le caratteristiche delle principali risorse monolingue di proverbi italiani oggi note e circolanti nella nostra Penisola.

Anzitutto, passiamo in rassegna qualche cifra: il dizionario di Guazzotti e Oddera (2006) riporta 11.000 proverbi, una quantità sicuramente elevata ma non insuperabile se pensiamo ai 20.000 detti proverbiali raccolti nell'*Atlante paremiologico italiano* a cura di Temistocle Franceschi ma anche al dizionario di Lapucci (2006), il quale presenta 25.000 proverbi. Come non citare, infine, il dizionario di Boggione e Massobrio (2004) con i suoi 30.000 proverbi *d'Italia*, e non *italiani*, come giustamente si sottolinea nella stessa introduzione all'opera (Boggione, 2004: XXIII), in quanto tale opera non raccoglie esclusivamente proverbi italiani ma anche proverbi dialettali provenienti da qualsiasi parte d'Italia.

Ci siamo ben presto resi conto, così, che il problema non stava nel reperimento di grandi quantità di proverbi in lingua italiana, bensì nell'aver un campione di proverbi qualitativamente adeguati per il nostro studio. Ecco perché una mera analisi quantitativa di queste opere non poteva, da sola, fornire abbastanza elementi per individuare quella più adatta alle nostre esigenze.

Anzitutto, non potevamo prendere in considerazione l'ultima opera citata, che è senz'ombra di dubbio quella più nutrita, ma è anche quella più forviante, in quanto, come ricordato, non prende solamente in considerazione i proverbi italiani, ma anche quelli dialettali, cosa che invece abbiamo volutamente escluso. Si noterà, specialmente nelle fasi successive della nostra analisi, che ciò che ci interessa maggiormente è lo studio del proverbio d'Italia e di lingua italiana e, allo stesso modo, il proverbio di Francia e di lingua francese. Tutta la sfera della paremiologia dialettale e regionale viene esclusa a priori per non creare

confusione né tantomeno distogliere l'attenzione dal nostro quesito fondamentale: come vengono utilizzati i proverbi italiani e i proverbi francesi in situazioni e contesti simili?

La scelta delle opere da seguire per formare la lista di proverbi da cercare in itWaK si è concentrata fatalmente sul dizionario di Guazzotti e Oddera (2006) e su quello di Lapucci (2006). Analizziamoli più nel dettaglio, facendo attenzione al materiale da essi presentato ma anche a come esso viene organizzato, nonché alle informazioni aggiuntive che vengono fornite. Entrambe le opere, così come indicano i rispettivi titoli, riguardano esclusivamente i proverbi italiani. Il dizionario di Guazzotti e Oddera (2006) presenta il materiale dapprima in ordine alfabetico (escludendo ai fini di tale ordinamento gli articoli con cui iniziano i proverbi e prendendo la prima parola utile), in seguito in ordine tematico (ovvero secondo macrocategorie tematiche, di volta in volta legate al significato paremiologico del proverbio o al suo significato compositivo⁴⁶, senza che sia chiaro a priori come l'utente debba utilizzare tale indice) e, infine, in ordine alfabetico secondo la parola chiave del proverbio, sebbene questa non sia sempre e facilmente individuabile (ad esempio, nulla vieta che la parola chiave individuata dagli autori non corrisponda alla parola chiave che l'utente che sta cercando il proverbio abbia in mente). Al contrario, il dizionario di Lapucci (2006) preferisce dapprima ordinare alfabeticamente i proverbi secondo la parola chiave di ogni entrata. Propone altresì un altro criterio di ricerca dei proverbi legato ad affinità concettuali: ciò significa che da un proverbio si può risalire a un altro proverbio presente nell'opera per affinità semantica. L'opera assegna, dunque, una grande importanza ai legami semantici che si possono costruire tra i diversi proverbi, i quali a loro volta costruiranno reti di sinonimie. Anche in questo caso, la parola chiave con cui si ordinano alfabeticamente i proverbi viene decisa dall'autore: se l'utente pensa a una parola chiave diversa da quella scelta nell'opera rischia di non ritrovare il proverbio desiderato. In più, diversamente dal dizionario di

⁴⁶ Ad esempio: *A caval donato non si guarda in bocca* si trova nella categoria "Animali" (Guazzotti e Oddera, 2006: 558). La categoria tematica in questo caso riguarda il significato compositivo e non paremiologico; non a caso, lo stesso dizionario definisce il proverbio nella maniera seguente: "un regalo deve essere accettato innanzitutto per le intenzioni del donatore, a prescindere dal valore e dall'importanza del dono stesso" (Guazzotti e Oddera, 2006: 21). Vediamo che nel significato del proverbio non si fa riferimento al fatto che il regalo debba essere un animale. Al contrario, il proverbio *Sbagliando s'impara* viene posto sotto la categoria "Errori e Imperfezioni" (Guazzotti e Oddera, 2006: 587). Se consideriamo il significato dato dallo stesso dizionario, ovvero "l'apprendimento avviene attraverso il riconoscimento e la presa di consapevolezza dei propri errori" (Guazzotti e Oddera, 2006: 461), notiamo come la categoria semantica faccia riferimento al significato paremiologico. Ora, la ragione di tale incertezza nel come si sono assegnate le categorie tematiche ai proverbi non viene chiarita nell'opera e possiamo solamente azzardare ipotesi legate, ad esempio, alla maggiore o minore vicinanza del significato paremiologico con quello compositivo. Qualunque sia la ragione, la conclusione che possiamo trarre è che queste categorie tematiche sono troppo incostanti da poter essere prese in considerazione per una ricerca scientifica sui proverbi.

Guazzotti e Oddera (2006), non si offre la possibilità di consultare un indice tematico, il quale potrebbe offrire un secondo aiuto all'utente, né tantomeno un indice meramente alfabetico dei proverbi. In fondo al dizionario si trova un indice analitico in cui dovrebbero essere inserite tutte le parole presenti nei proverbi del dizionario (in maniera da consentire il reperimento di un proverbio sulla base di una parola non considerata chiave), ma abbiamo notato che non vengono opportunamente indicate tutte le parole di ogni proverbio (ad esempio, *A caval donato non si guarda in bocca* presenta *cavallo* come parola chiave, ma non lo si riesce a reperire nell'indice analitico tramite la parola *bocca*).

Si noterà, dunque, che ogni dizionario presenta alcuni vantaggi per certi punti di vista e svantaggi per altri aspetti. Uno svantaggio comune ai dizionari presentati è, invece, la mancanza di informazioni legate all'uso dei proverbi. Effettivamente, si ha l'impressione che i dizionari menzionati, similmente a quanto avviene con altri dizionari di proverbi, siano tanto più zelanti nel definire ogni proverbio, nel fornirne esempi d'uso, tratti principalmente da fonti letterarie, nonché nel rintracciarne le origini, quanto meno attenti al loro reale utilizzo nell'italiano contemporaneo. In relazione ai tre dizionari citati, si trova un cenno alla frequenza d'uso (sebbene in maniera indiretta) solamente in Lapucci (2006): in tale opera si afferma che i proverbi vengono dapprima sistematizzati secondo le loro parole chiave e, successivamente, ordinati sotto la stessa parola chiave non alfabeticamente, "ma in base alla loro rilevanza" (Lapucci, 2006: IV). Tuttavia, non si specifica in che cosa consista la *rilevanza* di un proverbio, con quali criteri sia definita, come sia misurata, ma soprattutto non si assegna un grado di *rilevanza* a ogni proverbio. Facendo un esempio legato al proverbio *A caval donato non si guarda in bocca*, notiamo che sotto la sezione *Cavallo* vi è al primo posto il proverbio appena citato seguito, in ordine, da proverbi come *Se il cavallo è buono e bello, non guardar razza o mantello*; o ancora *Ognuno sa quanto corre il suo cavallo*; *Buon cavallo giunge e passa*; *Uomo a cavallo, sepoltura aperta*; *A cavallo da fieno, uomo di paglia*; etc. L'unica informazione riguardante la *rilevanza* di questi proverbi che possiamo dedurre è che *A caval donato non si guarda in bocca* è più *rilevante* degli altri proverbi citati, ma non vi è nulla che ci indichi quanto effettivamente essi vengano usati o in quali contesti. In ragione di tale approssimazione, non possiamo ritenere affidabile questa sorta di scala di frequenza dei proverbi.

2.9.2. Primo tentativo fallito

Pertanto, per compiere la nostra scelta ci siamo basati esclusivamente su mere ragioni di praticità, le quali ci hanno portato a prediligere il dizionario di Lapucci (2006). Tale dizionario, infatti, non ordinando alfabeticamente le proprie entrate, necessita di un metodo che faciliti il reperimento di un proverbio, specialmente se effettuato sulla base dell'elenco analitico posto in fondo all'opera. Questo consiste nell'assegnare un codice alfanumerico ai proverbi in base al principio seguente: come qualunque dizionario, anche questo è diviso per lettere, sotto le quali si trovano le parole chiave scelte per ogni proverbio che cominciano con la lettera corrispondente. Tutti i proverbi posti sotto ogni lettera hanno un numero progressivo, il quale sarà "1" per il primo, "2" per il secondo e così via fino all'ultimo proverbio classificato sotto la stessa lettera. La numerazione ricomincia così da "1" a ogni lettera. In questa maniera, l'indice analitico rimanda a ogni proverbio fornendone la lettera (quella con cui comincia la sua parola chiave) e un numero. Ad esempio, il proverbio *A caval donato non si guarda in bocca* (Lapucci, 2006: 217) ha come codice "C 1099". Ciò significa che è il proverbio numero millenovantanove della lettera C, in quanto la parola chiave è "Cavallo". Questo sistema ci è sembrato particolarmente utile al nostro scopo, poiché consente il reperimento casuale di una quantità prestabilita di proverbi tramite la generazione di numeri casuali (cfr. nota 47) corrispondenti ai proverbi da analizzare. Si trattava della sola metodologia possibile in quanto potevamo contare solo su risorse cartacee.

Per procedere alla selezione dei proverbi, abbiamo anzitutto calcolato la quantità di proverbi sotto ogni lettera dell'alfabeto (ricordiamo che la numerazione nel dizionario di Lapucci ricomincia ad ogni lettera). Così facendo, abbiamo appurato che i proverbi presenti nel dizionario sono in totale 25.178. Ci siamo fissati, come limite di proverbi da analizzare, la quantità di 500. Con qualche approssimazione, abbiamo stabilito che avremmo voluto analizzare il 2% della totalità dei proverbi. Per avere un campione che comprendesse tutte le lettere in cui è suddiviso il dizionario, abbiamo deciso di analizzare il 2% dei proverbi categorizzati sotto ogni lettera.

Nella tappa successiva, è stata generata una quantità sufficiente di numeri casuali⁴⁷ per ogni lettera, in maniera da individuare il campione del 2% di proverbi da analizzare.

In seguito, dopo aver generato tali codici, abbiamo riportato i proverbi corrispondenti in una tabella. La prima cosa che ci è balzata agli occhi era che i proverbi ottenuti erano molto

⁴⁷ La generazione dei numeri casuali è stata ottenuta tramite l'uso della funzione di generazione di numeri casuali di Microsoft Excel: la funzione usata, in linguaggio comprensibile a Excel, era "=CASUALE.TRA(X; Y)".

poco familiari a un italiano nativo contemporaneo. Sui 44 proverbi individuati sotto la lettera A, solo tre (*Gli affari sono affari; Si torna sempre ai vecchi amori; L'appetito vien mangiando*) potevano essere considerati proverbi noti, mentre un altro (*I primi amori sono i migliori*) poteva essere considerato la variante meno nota di un proverbio ben più conosciuto (*Il primo amore non si scorda mai*) posto immediatamente prima a quello selezionato e quindi escluso a priori dalla lista di proverbi casuali. Per avvalorare le nostre intuizioni, abbiamo cercato i rimanenti proverbi prima in itWaK, ma anche in Internet per ampliare la ricerca, e abbiamo notato che nella stragrande maggioranza dei casi essi non sono attestati oppure lo sono ma solamente in siti Internet che si occupano di proverbi. Ciò significa che in media solo un proverbio su quindici individuati secondo questa modalità è attestato in internet. Certo, è errato ritenere che i proverbi non attestati non siano utilizzati nell'italiano contemporaneo (ricordiamo che ogni osservazione dedotta da un corpus o dalla Rete è pur sempre valida per il campione di dati considerato, che può certamente essere rappresentativo, ma non sarà mai lo specchio esatto della realtà). Ciononostante, si può affermare che il loro uso è molto probabilmente marginale rispetto a quello dei proverbi che trovano riscontro in itWaK o in Rete. Inoltre, considerando che per portare a termine la nostra ricerca avremmo dovuto utilizzare itWak e frWak, era essenziale che il nostro materiale fosse presente in questi due corpora. Pertanto, il primo tentativo di reperimento di proverbi nei corpora era fallito.

2.9.3. Marcatori di proverbio e secondo tentativo

In base alle nostre considerazioni iniziali (cfr. § 2.8 e § 2.9) e riflettendo sulle ragioni del nostro primo fallimento, abbiamo compreso che l'unica strada percorribile poteva essere legata all'approccio *corpus-driven*.

Per tentare di reperire il nostro campione di proverbi secondo questa modalità, abbiamo preso spunto dall'area di ricerca che si trova al confine tra lo studio delle espressioni idiomatiche e la linguistica dei corpora. Abbiamo, cioè, ripreso la nozione di *marcatori di metafora* (cfr. § 2.8) per applicarla ai proverbi, al fine di considerare l'ipotesi dell'esistenza di *marcatori di proverbio*.

Similmente a quanto accade, in generale, per la linguistica del proverbio e, in particolare, per gli studi conversazionali riguardanti il proverbio, i lavori sui *marcatori di proverbio*, ovvero sulle formule che introducono od occorrono frequentemente con tali enunciati, non sono molto avanzati. In letteratura, si trovano informazioni su questi elementi, sebbene vi sia una carenza, tutt'altro che trascurabile, di dati quantitativi e qualitativi

derivanti da corpora linguistici. Gli esempi riportati sono, infatti, frequentemente episodici, occasionali, individuati ad hoc e più raramente si presentano dati estrapolati da banche dati di qualsiasi tipo con le relative statistiche.

Tra le principali ricerche riguardanti le formule che solitamente co-occorrono con l'enunciazione dei proverbi ricordiamo quelle elaborate da Schapira (2000), ma anche da Kleiber (1999) e da Cram (1983). Ciò che emerge nei lavori dei tre studiosi citati è la possibilità che il proverbio venga accompagnato da formule che ne indichino lo status paremiologico, sebbene questo non sia indispensabile, come emerge in maniera trasversale dai lavori dei tre autori citati. Schapira (2000: 89-90) sostiene che certamente esistono formule del tipo "*comme on dit*", "*comme dit le proverbe*" che introducono il proverbio e volte a sottolineare che l'enunciato a cui si fa riferimento è caratterizzato da uno status citazionale e non è un'invenzione libera del parlante o dello scrivente. Tuttavia, la stessa Schapira sostiene che tale procedimento era particolarmente frequente nel '600, mentre dalla fine di quel secolo fino ai nostri giorni il proverbio ha cominciato a inserirsi nel discorso senza essere accompagnato da alcuna formula, in quanto è la sua stessa notorietà ad assicurarne il riconoscimento da parte dell'interlocutore. Tuttavia, approfondendo il discorso dei vincoli linguistici necessari perché un proverbio sia utilizzato all'interno di realtà testuali, Schapira (2000: 91-92) mette in luce dapprima che i proverbi si presentano per forza come citazioni e, dunque, possono essere introdotti da qualsiasi proposizione, comprese quelle proposizioni in cui compaiono verbi di opinione alla prima persona singolare (cosa che non sarebbe possibile in teoria, ma anche in pratica, in quanto i proverbi, essendo assimilati a luoghi comuni, leggi umane e stereotipi, non possono essere mostrati come pensieri propri). Approfondendo l'analisi, si scopre che, da uno studio condotto dalla studiosa su un corpus di articoli di giornali francesi (specialmente *L'Express* e *Le Nouvel Observateur*) su un periodo di cinque anni (1993-1998), sono emerse essenzialmente due modalità di inserimento del proverbio in contesto: l'inserimento senza alcuna formula introduttiva e l'inserimento con una formula che mette in evidenza il carattere di enunciato di notorietà generale del proverbio, come "*on le sait*"; "*jusqu'à en oublier que*"; "*la bonne sagesse populaire*" (Schapira, 2000: 92).

Kleiber (1999) considera la questione indirettamente e di riflesso, in quanto è maggiormente interessato a studiare il grado di coinvolgimento del locutore nell'enunciazione del proverbio, ma coglie l'occasione per analizzare le modalità di inserimento del proverbio in

contesto. In particolare, si concentra sulle modalità con le quali si introduce solitamente un giudizio collettivo (inteso in opposizione a un giudizio individuale, ovvero un giudizio espresso tramite un enunciato il cui autore è colui che lo emette). L'autore ribadisce a più riprese il fatto che il proverbio sia un enunciato ecoico in quanto fa eco a un'affermazione o a un pensiero di altri, cosa che rende indiscutibile il fatto che l'autore del proverbio non corrisponda all'enunciatore. Il parlante non sarà dunque responsabile della forma del proverbio poiché esso è riportato come se fosse una citazione. Da ciò scaturiscono diverse considerazioni legate segnatamente alle espressioni metalinguistiche che accompagnano frequentemente i proverbi, che possono essere "*comme on dit*", in opposizione alla tipica espressione che accompagna le classiche citazioni del tipo "*comme (le) dit X*" in cui X rappresenta un individuo ben preciso; "*on a bien raison de dire que...*" o "*si j'en crois la sagesse populaire*" o ancora "*si j'en crois la sagesse des nations*"; "*comme (le) dit le proverbe*" o "*comme dit un proverbe*" (Kleiber, 1999: 58). Si nota come le stesse espressioni introduttive siano più difficilmente accettabili al passato, come mostrato nei seguenti casi: *"*Comme on a dit*" seguito da proverbio oppure *"*Comme l'a dit le proverbe*", in quanto queste espressioni fanno riferimento a una lettura episodica dell'enunciazione del proverbio che contrasterebbe con il carattere collettivo del giudizio espresso dal proverbio (Kleiber, 1999-59). Tutto ciò dimostra quanto sia in fondo poco importante quale sia espressione che introduce un proverbio, poiché qualsiasi espressione potrebbe essere in grado di introdurre un proverbio a condizione che non si neghi o non si metta in dubbio il carattere non individuale del proverbio. Ecco perché sono accettabili espressioni introduttive in cui figura un enunciatore, che non sarà per definizione l'enunciatore primo del proverbio, bensì colui che ribadisce in primis la validità generale del proverbio e che sottolinea quanto essa sia applicabile alle situazioni particolari (Kleiber, 1999: 60-68). Tutto quanto fin qui espresso mostra come sia tutt'altro che facile reperire un marcatore di proverbio abbastanza generale e trasversale che possa fungere da denominatore comune alle molteplici e imprevedibili espressioni introduttive del proverbio.

Tuttavia, i vantaggi legati all'individuazione dei *marcatori di proverbio* come porta di accesso al reperimento dei proverbi stessi sono molti e di grande rilevanza, in quanto maggiormente legati all'approccio *corpus-driven*. Come succede generalmente con questo metodo, anche utilizzando i *marcatori di proverbio* il paremiologo, prima di analizzare il materiale a sua disposizione, interroga il corpus, dandogli la possibilità di "esprimersi da solo"

e di restituire una gamma di proverbi non noti a priori, ma rispondenti a certi criteri formali prestabiliti.

Oltre a tutto ciò, il nostro tentativo era volto a reperire il maggior numero di proverbi esistenti nel corpus, evitando però di incorrere nella trappola costituita da altri tipi di frasi o espressioni idiomatiche che non fossero proverbi ma semplicemente elementi appartenenti al campo gnomico o idiomatico.

Siamo dunque giunti a una soluzione, che probabilmente non è la migliore, ma è stata sicuramente l'unica a poter rispondere a tutte le nostre necessità: trovare i proverbi contenuti nel corpus itWaC, partendo dalla concordanza dalla parola *proverbio*. In altre parole, cercando la parola *proverbio* avremmo probabilmente potuto ritrovare in primis quelle espressioni in cui si utilizzava tale parola, con la speranza che essa fosse impiegata con funzione di segnalazione della presenza di un proverbio enunciato nella porzione di testo immediatamente precedente o successiva.

I primi risultati ottenuti sono stati quantitativamente e qualitativamente interessanti: abbiamo, infatti, trovato ben 3.706 occorrenze della parola *proverbio*, che corrispondono a circa 1,9 pmw (*per million words*) del corpus. Da una prima sommaria analisi, si è riscontrata una notevole presenza di proverbi effettivamente in uso oggi, ma anche di proverbi stranieri, compresi proverbi italiani ma enunciati in dialetti e lingue locali, e di proverbi appartenenti a opere letterarie anteriori al '900. Inoltre, si sono da subito notati casi, sebbene non molto numerosi, di ritrovamento di proverbi avulsi da qualsiasi contesto enunciativo in quanto enunciati in elenco semplicemente per il piacere di ricordare o enumerare proverbi. Ciononostante, gli esempi d'uso dei proverbi all'interno di testi coerenti sono apparsi da subito molti e molto interessanti.

Prima di cominciare a mettere mano ai risultati, abbiamo deciso di condurre una ricerca simile anche per il corpus francese. Pertanto abbiamo scelto di cercare le concordanze della parola *proverbe* all'interno del corpus frWaC. I risultati sono sembrati da subito molto incoraggianti, avendo trovato 3.537 occorrenze della parola, ossia 2,2 pmw del corpus. Gli stessi inconvenienti riscontrati per il corpus itWaC sono emersi da subito anche nel caso di frWaC: apparivano, infatti, molte occorrenze di proverbi stranieri o dialettali, proverbi utilizzati in opere letterarie antiche o proverbi facenti parte di siti particolarmente votati alla paremiologia o all'interesse storico, folkloristico o culturale legato ai proverbi. Tuttavia,

anche nel caso del francese, sono emersi da subito numerosi e interessanti esempi di proverbi utilizzati in testi coerenti.

Per tutte queste ragioni, abbiamo deciso di seguire questa strada di ricerca, sebbene avessimo considerato da subito i pro e i contro di tale approccio. I pro riguardavano tutto quanto già espresso relativamente ai vantaggi dell'utilizzo di un procedimento *corpus-driven*. I contro, invece, consistevano nella non esaustività dei risultati reperiti: come mostrato, tra gli altri, dai tre autori citati in questo § 2.9.3 (Kleiber, 1999; Shapira, 2000; Cram, 1983), l'enunciazione del proverbio non viene sempre accompagnata da una formula che ne indichi lo status. In altri termini, anche in ragione del fatto che il proverbio è una frase e non un enunciato, si possono trovare proverbi all'interno di testi senza che nulla, a livello metalinguistico, ne indichi la presenza. Oltre a ciò, come si vedrà nella fase di analisi dei proverbi reperiti nei due corpora, non è sempre detto che quanto i parlanti segnalano come proverbio sia effettivamente un proverbio (cfr. § 4.5). Tali osservazioni trovano una prima conferma in Moon (1998: 47), in cui si nota che il corpus rimette in discussione la netta separazione tra espressione idiomatica (il che può essere esteso, a nostro avviso, all'ambito paremiologico) e altre tipologie di elementi linguistici.

In questa bilancia di pro e contro, i vantaggi superavano gli svantaggi o li rendevano, per lo meno, più tollerabili. Per tale ragione, e nonostante i limiti evidenziati, abbiamo deciso comunque di percorrere questa strada, da noi giudicata fruttuosa fin dall'inizio.

In conclusione, abbiamo tentato di mostrare, nel presente capitolo, le potenzialità della linguistica dei corpora e, segnatamente, le ragioni per cui prediligere l'approccio *corpus-driven* allo studio dei proverbi. Tale approccio è l'unico in grado di prendere in considerazione un insieme di dati non definiti a priori. In maniera particolare, se l'oggetto di studio è il proverbio, l'approccio *corpus-driven* è il solo che garantisca una ricerca esaustiva sia quantitativamente sia qualitativamente, poiché non tiene conto della variabilità, alquanto elevata nel caso dei proverbi.

Tuttavia, questo *modus operandi* può risultare altrettanto complicato, poiché non vi è nulla, a livello di annotazione morfosintattica o di elementi lessicali e sintattici, che indichi con precisione che una determinata frase è un proverbio. Si è, infatti, notato che i proverbi si comportano formalmente come frasi libere. L'unico elemento formale che può aiutare a reperire automaticamente proverbi all'interno di un grande corpus è l'etichetta *proverbio* che viene spesso impiegata dall'enunciatore per segnalare lo status paremiologico di quanto ha

enunciato o sta per enunciare. Sebbene anche questo metodo sia lungi dall'essere perfetto, abbiamo deciso di applicarlo, per cercare e, successivamente, analizzare proverbi ed eventuali *pattern* ricorrenti che possano essere rilevanti a tale analisi.

Facendo una breve ricerca esplorativa, abbiamo osservato che uno studio di questo tipo può portare a risultati interessanti e comparabili se effettuato su due corpora di riferimento molto grandi come possono essere itWaC e frWaC.

Esporrremo i dettagli di questa ricerca nel prossimo capitolo, assieme ai principali dati ottenuti. Il nostro obiettivo, lo ricordiamo, rimane quello di proporre strumenti concreti che possano essere utili a professionisti delle lingue, come interpreti o traduttori, nell'ardua ricerca di corrispondenze tra proverbi in L1 e proverbi in L2.

3. Estrazione automatica di proverbi italiani e francesi

3.1. Percorso seguito nell'individuazione dei proverbi da analizzare

Nel capitolo precedente, sono state illustrate alcune delle modalità possibili per individuare automaticamente proverbi in grandi corpora linguistici. Come già osservato nello stesso capitolo, non sempre le modalità più diffuse o più utilizzate in altri ambiti possono essere applicate ai proverbi, in ragione della loro specificità. Si è pertanto deciso di reperire proverbi sulla base di quelli che abbiamo definito *marcatori di proverbio*, tentando di individuare quel marcatore più frequente che possa parallelamente garantire il ritrovamento del numero più alto di proverbi. In base ai nostri tentativi, tale marcatore altro non è che la stessa parola *proverbio*⁴⁸.

Abbiamo così eseguito la concordanza di questa parola. Per fare ciò, ci siamo serviti del programma di gestione dei corpora WaCky, ossia NoSketch Engine⁴⁹. Vediamo nel dettaglio i passaggi necessari per ottenere le concordanze sia della parola *proverbio* sia della parola *proverbe*.

Dall'interfaccia di ricerca del programma di NoSketch Engine, si è anzitutto selezionato il campo CQL, acronimo di Corpus Query Language (cfr. fig. 1). A questo indirizzo, <<https://www.sketchengine.co.uk/corpus-querying/>>, è possibile ritrovare alcune indicazioni utili per effettuare ricerche semplici e complesse. Selezionando “word” dal menù a tendina in basso alla pagina recante l'etichetta “Default attribute”, è possibile cercare una parola precisa semplicemente virgolettando la parola stessa. Abbiamo così digitato i caratteri “*proverbio*” nella maschera di ricerca in basso chiamata “CQL”. Per maggiore praticità, si indica l'URL esatto della pagina relativa a NoSketch Engine con già selezionato il corpus itWaC:

```
<http://nl.ijs.si/noske/wacs.cgi/first?corpname=itwac&reload=1&iquery=&queryselector=wordrow&lemma=&lpos=&phrase=&word=&wpos=&char=&cql=&default\_attr=word&fc\_lemword\_window\_type=both&fc\_lemword\_wsize=5&fc\_lemword=&fc\_lemword\_type=all&fc\_pos\_window\_type=both>.
```

⁴⁸ Ci siamo attenuti strettamente alla parola *proverbio* al singolare, evitando di proposito il plurale. Questo perché, in un saggio esplorativo eseguito nella fase di ricerca dei marcatori di proverbi, si è notato che la parola *proverbi* raramente accompagna l'enunciazione di un proverbio e più frequentemente si riferisce ai proverbi in generale.

⁴⁹ URL: <<http://nl.ijs.si/noske/index-en.html>>. Ultima visita: 28 maggio 2016. Ricordiamo, altresì, che il programma è fruibile da parte di tutti gratuitamente e non è necessario scaricarlo: è sufficiente avere una connessione Internet attiva per poter lavorare con questo programma da qualsiasi computer.

Osservando la fig. 1, si noterà che la parola proverbio è scritta tra virgolette: questo perché, avendo selezionato “word” come attributo di default, è sufficiente digitare ogni parola desiderata tra virgolette perché il programma esegua le concordanze di ognuna delle parole tra virgolette. Altrimenti, per rispettare la sintassi CQL, se non avessimo impostato “word” come attributo di default, avremmo dovuto digitare la seguente stringa: [word=“proverbio”].

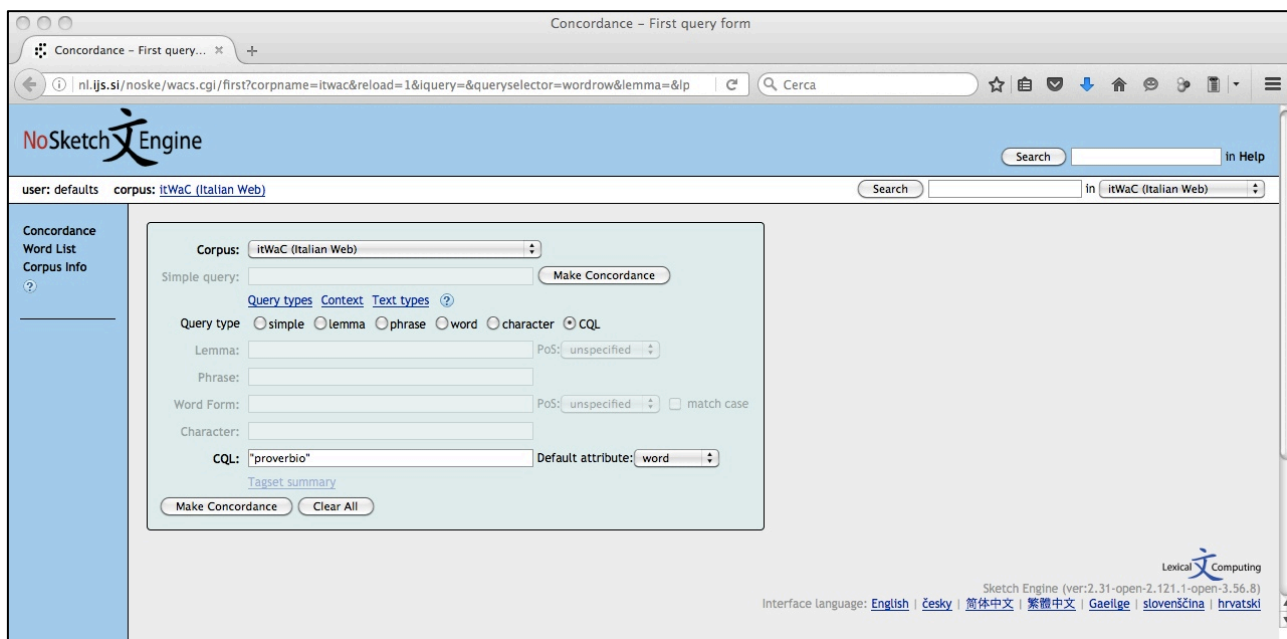


Fig. 1. Maschera di ricerca del programma NoSketch Engine con il corpus itWaC selezionato.

Abbiamo così ottenuto 3.706 risultati, ovvero 3.706 casi in cui viene impiegata la parola *proverbio* (fig. 2). Per nostra praticità, si è deciso di ordinare i risultati semplicemente secondo la prima parola a destra del nodo: questo ordine può essere particolarmente utile per disporre i duplicati in posizioni contigue, al fine di renderne più facile l’identificazione.

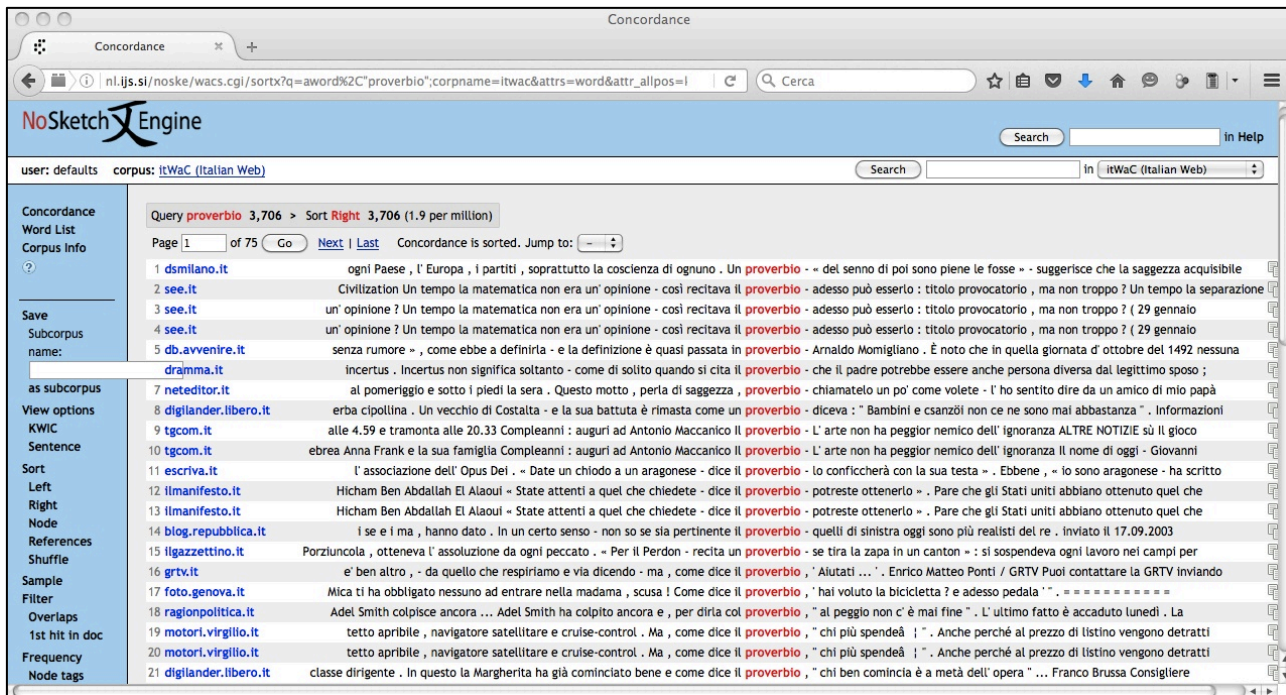


Fig. 2. Concordanza della parola proverbio.

Il primo aspetto che possiamo notare è che, in realtà, l'ordine dei risultati non è di grande importanza: NoSketch Engine considera i segni di punteggiatura alla stregua delle parole. Per questa ragione, incontreremo dapprima gli esempi in cui la parola *proverbio* è seguita dal segno "-", poi dal segno ",", e così via secondo il seguente ordine: ";", ":", "!", "?", "/", ".", "...", " ", "(", ")", "[", "*", "\". Pertanto, ordinare i risultati secondo il contesto a destra può essere utile, ma ha comunque il difetto di rendere necessaria una lettura attenta di tutti i risultati da parte del ricercatore.

Successivamente, abbiamo selezionato alcune impostazioni di visualizzazione. In primo luogo, abbiamo deciso che fosse necessario visualizzare l'URL di ogni risultato⁵⁰: questo perché speravamo, in caso di bisogno, di poter accedere direttamente alle pagine in cui era stata utilizzata la parola *proverbio*. Poter leggere il contesto in cui occorrevano i diversi esempi ottenuti sarebbe stato di grande aiuto per farci capire la validità o meno dei risultati stessi (nel prossimo § 3.2 si spiegherà meglio questo punto, fornendo maggiori dettagli circa la necessità di scartare certi risultati); tuttavia, una quantità importante dei link forniti dal

⁵⁰ In Fig. 2 si è deciso di mostrare non l'URL di ogni concordanza, bensì solamente il dominio: la scelta è dovuta a mere ragioni di praticità, in quanto gli URL, essendo composti da moltissimi caratteri, sono più lunghi di una schermata. Così facendo, correavamo il rischio di fare un'istantanea di una finestra in cui si osservava solo un elenco di URL. Per questa ragione, abbiamo scelto di visualizzare, per l'immagine di Fig. 2, solamente i domini.

corpus non sono più esistenti. Tale considerazione ha imposto un cambiamento di strategia: per considerare a dovere la validità di ogni risultato, potevamo fare affidamento solamente sul cotesto restituito dal corpus. Pertanto, abbiamo aumentato i caratteri visualizzati per ogni risultato da 100 a 250: 250 erano dunque i caratteri sia a sinistra del nodo e sia a destra del nodo. Se per alcune concordanze il cotesto così ricavato era persino troppo ampio, per altre invece è stato essenziale.

3.2. Vaglio dei risultati

Dopo aver eseguito le ricerche della parola *proverbio* e *proverbe*, si è resa necessaria una fase di pulitura dei risultati ottenuti, perché non tutte le volte che si usano queste due parole nelle rispettive lingue si indica un proverbio. Questo punto è di grande rilevanza: l'approccio *corpus-driven* che ci siamo proposti di utilizzare è un approccio che permette il reperimento automatico di proverbi che condividono la caratteristica di essere stati definiti tali dall'autore del testo in cui compaiono. Tuttavia, non esistendo alcuna maniera per riconoscere automaticamente e in maniera affidabile i proverbi, bisognava passare al setaccio i risultati ottenuti in maniera manuale. Prima di capire come procedere, abbiamo deciso di analizzare i primi 150 risultati del corpus itWaC per vedere quali di questi era necessario scartare e quali si potevano convalidare. L'obiettivo, inoltre, consisteva nel reperire una metodologia chiara e applicabile anche al corpus frWaC per compiere lo stesso processo di scrematura, cosa che avrebbe garantito l'ottenimento di dati comparabili.

3.2.1. “Proverbi non italiani”, “proverbi non francesi”

Analizzando i primi 150 risultati, abbiamo notato che vi erano molti proverbi di origine africana e cinese, oltre a parecchi proverbi d'Italia ma dialettali. Abbiamo così capito, fin da subito, che avremmo dovuto affrontare il problema della presenza di proverbi non italiani (in lingua originale e tradotti in italiano) nel nostro corpus. Avrebbe avuto senso prendere in esame anche i proverbi stranieri e dialettali? Considerando che il nostro obiettivo consisteva nell'analisi dei proverbi italiani (e in italiano), abbiamo deciso che questi risultati non fossero rilevanti ai fini del nostro studio. Inoltre, avevamo notato che gli utenti della Rete decidono spesso di citare proverbi stranieri più per una ricerca di un gusto esotico che per un vero interesse comunicativo. Lo stesso vale anche per i proverbi locali e dialettali, che vengono citati in numerose occasioni soltanto per rievocare suggestioni di epoche passate o analizzare usi e costumi del passato.

Le regole che abbiamo stabilito per individuare i proverbi stranieri erano le seguenti: se la parola *proverbio* fosse stata seguita da un aggettivo di nazionalità o popolazione, da aggettivi come “locale”, “regionale”, “dialettale” oppure dal nome di città o di stato, allora il proverbio sarebbe stato da considerarsi un “proverbio non italiano” (etichetta scelta per comprendere anche quei risultati relativi a proverbi provenienti dall’Italia ma dialettali o locali); inoltre, avremmo anche proceduto all’eliminazione di quei risultati in cui fosse emerso chiaramente dall’URL la provenienza estera o locale dei proverbi.

Le stesse regole sono state riapplicate alla parte riguardante il corpus francese. La sensazione che si è avuta da subito è che i proverbi non francesi ma di Francia fossero meno numerosi rispetto ai proverbi non italiani ma d’Italia, mentre aumentavano a dismisura i risultati di proverbi non francesi ma di lingua francese di altri paesi e, specialmente, delle ex colonie. Nonostante queste prime differenze osservate, benché quantificate solo una volta conclusa la fase di scrematura di tutte le concordanze, le considerazioni sarebbero dovute rimanere le stesse anche per il corpus francese. Pertanto, come abbiamo deciso di eliminare tutti i risultati del corpus italiano relativi a localismi, regionalismi o forestierismi, allo stesso modo abbiamo scartato i proverbi non francesi, oppure quelli in lingua francese ma provenienti da altri paesi o, ancora, i proverbi in francese ma seguiti (e raramente preceduti) dal nome di città, località, dipartimenti o regioni francesi.

3.2.2. Considerazioni sui “proverbi non italiani / non francesi”

Siamo riusciti a identificare in maniera semiautomatica i “proverbi non italiani” e i “proverbi non francesi” grazie alla presenza di aggettivi, nomi di paesi, di regioni o di località indicanti la provenienza dei proverbi. Questo sistema permette anche l’individuazione semiautomatica delle origini dei proverbi utilizzati dagli utenti della Rete, sia italiani sia francesi. Basta, infatti, creare un secondo corpus di soli esempi contenenti proverbi stranieri e ordinare alfabeticamente i risultati secondo la prima parola a destra del nodo (che nella quasi totalità dei casi corrisponderà all’aggettivo di provenienza) per evidenziare i gruppi di proverbi aventi la stessa provenienza e calcolare, appunto, le diverse origini degli stessi. In Fig. 1, osserviamo un esempio di concordanza della parola *proverbio* all’interno del nuovo corpus creato a partire dagli esempi scartati nella fase di pulitura del corpus itWaC. Per eseguire questo compito, abbiamo utilizzato il software di gestione di corpora AntConc 3.2.4m per Macintosh OS X 2011.

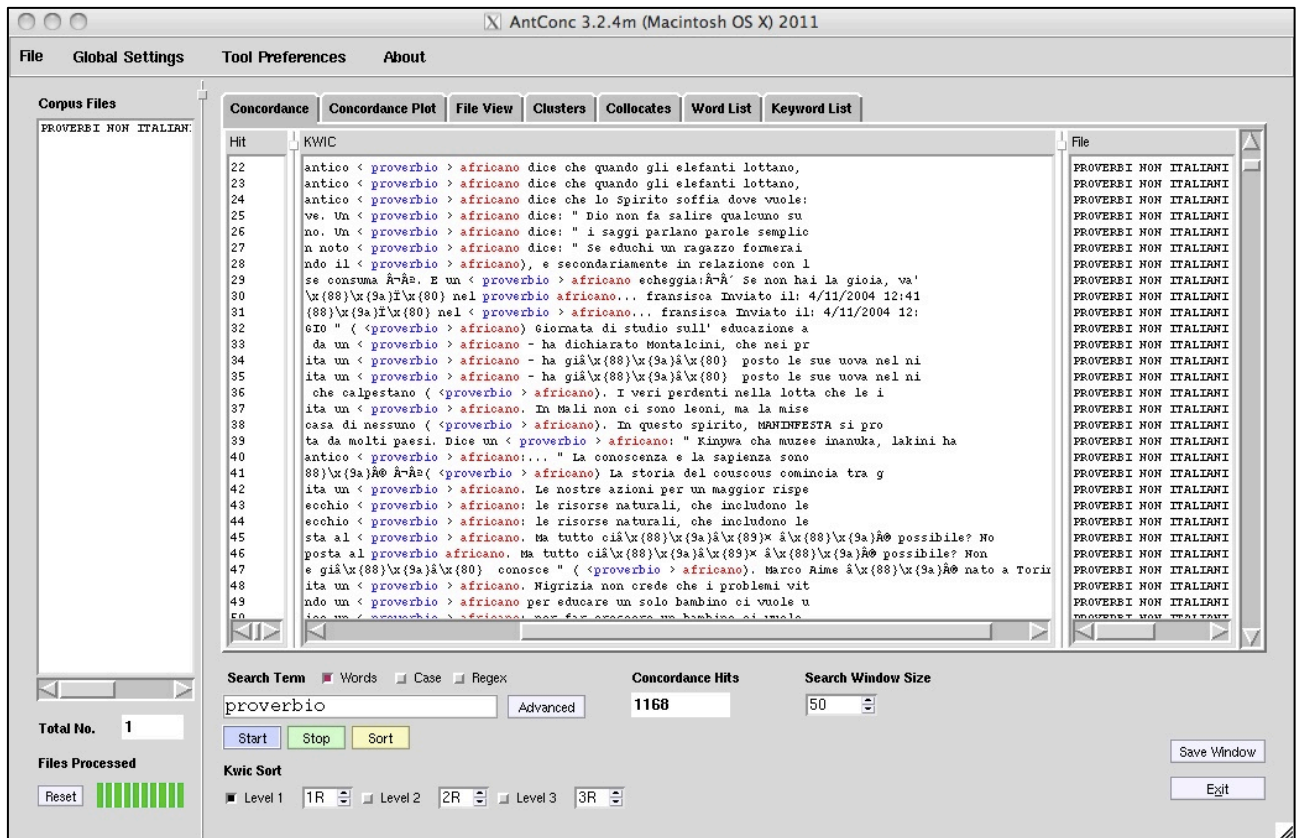


Fig. 3. Concordanza di *proverbio* eseguita con il software di gestione di corpora AntConc 3.2.4m utilizzando il corpus costruito sulla base dei risultati eliminati dal corpus itWaC e relativi agli esempi di “proverbi non italiani”.

Sebbene in questa prima parte del capitolo si stiano ancora esaminando le prime tappe raggiunte per individuare la metodologia di scrematura delle concordanze di *proverbio* e *proverbe*, presentiamo i dati ottenuti una volta conclusa l'intera fase di analisi e pulitura delle concordanze. I casi di cui ci occuperemo ora riguardano quegli esempi in cui è stato possibile capire che i proverbi non potevano essere considerati italiani in quanto o di provenienza estera oppure di estrazione dialettale o regionale, il che è possibile, nella stragrande maggioranza dei casi, grazie alla presenza di un aggettivo di provenienza, di nomi di paesi o di località che seguono il sostantivo *proverbio*. Ricordiamo, infine, che lo stesso procedimento è stato utilizzato per le concordanze della parola *proverbe* ottenute sulla base del corpus frWaC. È interessante vedere quali sono i proverbi stranieri maggiormente citati nelle diverse lingue, non tanto per analizzare i proverbi in sé, quanto piuttosto per vedere da quali bacini attingano le due lingue analizzate per reperire e utilizzare “proverbi non italiani” o “non francesi”.

Questo potrebbe essere un indice di affinità, vicinanza o anche solo di elevato grado di contatti con le culture che hanno originato i proverbi impiegati.

Anzitutto, concentriamoci sulle quantità totali dei proverbi stranieri individuati a partire dalle concordanze di *proverbio* e *proverbe*. Si vedrà più avanti (cfr. § 3.3) che, in generale, si usano più proverbi stranieri in francese che in italiano, almeno nei casi in cui si segnalano tali proverbi con le parole *proverbio* e *proverbe*. Per la precisione, sono stati riscontrati 848 proverbi non francesi tra le concordanze di *proverbe* e 637 proverbi non italiani tra le concordanze di *proverbio*. In percentuale, considerando che il numero di “proverbi non francesi” è più elevato rispetto a quello dei “proverbi non italiani” e che le concordanze del corpus frWaC sono meno numerose rispetto a quelle di itWaC, si osserva che i “proverbi non francesi” sono il 24% della totalità delle concordanze di frWaC, una percentuale sicuramente più elevata di quella dei “proverbi non italiani”, pari al 17% delle concordanze di itWaC.

Nelle prossime due tabelle (Fig. 4 e Fig. 5) si riportano tutte le provenienze dei proverbi non italiani e non francesi presenti tra le concordanze di *proverbio* e *proverbe*, con l’indicazione del numero di volte in cui troviamo la stessa origine utilizzata per proverbi diversi (da notare che, prima di procedere a questo calcolo, sono stati eliminati i duplicati di risultati).

itWaC	
Provenienza	Numero di occorrenze
Cinese	143
Africano	50
Arabo	43
Siciliano	25
Russo	23
Latino	22
Indiano	21
Giapponese	19
Inglese	18
Tedesco	18
Orientale	17
Polacco	16
Francese	15

Toscano	14
Locale	10
Greco	9
Spagnolo	9
Turco	9
Ebraico	8
Napoletano	7
Tibetano	7
Tuareg	7
Irlandese	6
Veneto	6
Yiddish	6
Americano	4
Asiatico	4
Genovese	4
Latino-americano	4
Malgascio	4
Messicano	4
Milanese	4
Argentino	3
Croato	3
Magrebino	3
Partenopeo	3
Persiano	3
Romano	3
Scozzese	3
Senegalese	3
Zen	3
Brasiliano	2
Calabrese	2
Creolo	2
Egiziano	2
Emiliano	2
Lombardo	2
Malese	2
Maori	2
Meridionale	2
Piemontese	2
Sardo	2
Serbo	2
Veneziano	2
Vietnamita	2
Abruzzese	1

Afghano	1
Amerindiano	1
Andaluso	1
Armeno	1
Bolognese	1
Bresciano	1
Britannico	1
Camerunese	1
Canadese	1
Cubano	1
Estone	1
Finlandese	1
Germanico	1
Indigeno	1
Malayalam	1
Peruviano	1
Portoghese	1
Reggiano	1
Romagnolo	1
Rwandese	1
Scandinavo	1
Somalo	1
Tunisino	1
Yemenita	1
Zingaro	1

Fig. 4. Lista delle provenienze dei “proverbi non italiani” in ordine decrescente in base alle relative occorrenze. Tali provenienze sono state copiate e incollate senza l’apporto di alcuna modifica se non l’indicazione del nome di località prima delle eventuali preposizioni, che sono state indicate tra parentesi.

frWaC	
Provenienza	Numero di occorrenze
Chinois	244
Africain	92
Arabe	60
Indien	22
Russe	20
Allemand	18
Japonais	17

Anglais	16
Espagnol	16
Corse	15
Antillais	13
Latin	13
Persan	13
Tuscarora	12
Breton	10
Local	9
Grecs	8
Kabyle	8
Italien	7
Vietnamien	7
Amazigh	6
Américain	6
Juif	6
Tibétain	5
Algérien	4
Alsacien	4
Danois	4
Israël (dans le pays d')	4
Hindou	4
Hongrois	4
Berbère	3
Créole	3
Wolof (en)	3
Haïtien	3
Malgache	3
Masaï	3
Polonais	3
Populaire	3
Québécois	3
Romain	3
Rwandais	3
Yoruba	3
Afghan	2
Alpin	2
Ardechois	2
Asiatique	2
Bantou	2
Bouddhiste	2
Brésilien	2

Burundais	2
Béarnais	2
Catalan	2
Congolais	2
Coréen	2
De chez moi	2
De ma mère	1
De ma région	1
Du Médoc	2
Finlandais	2
Groenlandais	2
Hébreu	2
Irlandais	2
Islandais	2
Marocain	2
Occitan	2
Oriental	2
Portugais	2
Provençal	2
Roumain	2
Scandinave	2
Serbe	2
Sioux	2
Slovaque	2
Suédois	2
Touareg	2
Turc	2
Zaïrois	2
Zoulou	2
Amérindien	1
Ashanti	1
Augeron	1
Balkanique	1
Bambara	1
Basque	1
Belge	1
Californien	1
Castillan	1
Champenois	1
Cherokee	1
Chiite	1
Colombien	1
Comorien	1

Croate	1
Côte d'Ivoire	1
Rome Antique (de la)	1
De mon pays	1
De sa culture	1
Mali (des Bwa du)	1
Indiens Cree du Canada	1
Druze	1
Ali (du calife)	1
Du Cheshire	1
Étranger	1
Finnois	1
Flamand	1
Gallois	1
Gaulois	1
Gitan	1
Hassaniya	1
Haut savoyard	1
Indigène	1
Inuit	1
Kongo	1
Lacédémonien	1
Lao	1
Lavallois	1
Levantin	1
Machal	1
Malinka	1
Marseillais	1
Mayenneux	1
Mbochi	1
Mexicain	1
Montpéliairain	1
Navajo	1
Nordiste	1
Normand	1
Néo-zélandais	1
Népalais	1
Palestinien	1
Parisien	1
Peul	1
Picard	1

Polien	1
Péruvien	1
Sanskrit	1
Savoyard	1
Sibérien	1
Somalien	1
Suisse	1
Swahili	1
Talmudique	1
Tchèque	1
Togolais	1
Tootsie	1
Nice (trivial de)	1
Troyen	1
Tunisien	1
Vosgien	1
Védique	1
Wallon	1
Yi	1
Yiddish	1

Fig. 5. Lista delle provenienze dei “proverbi non francesi” in ordine decrescente in base alle relative occorrenze. Tali provenienze sono state copiate e incollate senza l’apporto di alcuna modifica se non l’indicazione del nome di località prima delle eventuali preposizioni, che sono state indicate tra parentesi.

Più nel dettaglio, notiamo che le prime provenienze per numero di occorrenze sono molto simili tra i proverbi stranieri di itWaC e quelli di frWaC: i proverbi cinesi sono i più frequenti per itWaC con 143 occorrenze contro le 244 di frWaC. La seconda provenienza per frequenza è, per entrambi i corpora, l’Africa, con 50 occorrenze per itWaC e 92 per frWaC. La terza provenienza è indicata dall’aggettivo “arabo” e “arabe”, che totalizzano rispettivamente 43 e 60 occorrenze.

Considerando la notevole quantità di provenienze diverse (ne abbiamo contate 81 in italiano contro 153 in francese), e al fine di rendere più chiari i risultati, abbiamo deciso di suddividere le provenienze per continenti o macro-aree, come Europa, Europa antica (Roma antica, Grecia antica, i celti, i galli, ecc.), Eurasia, Asia, Medio-Oriente, Africa, America e Oceania, oltre a Italia (per i proverbi regionali o dialettali di itWaC) e Francia (per i medesimi proverbi relativi al corpus frWaC). Vediamo nei grafici di Fig. 6 e Fig. 7 la distribuzione dei proverbi per macro-aree geografiche.

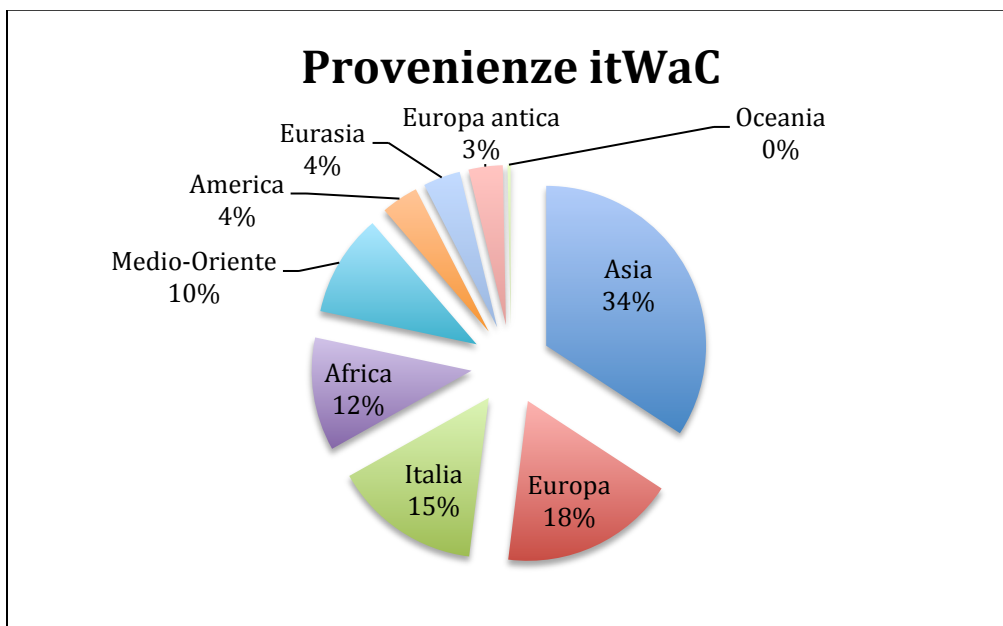


Fig. 6. Distribuzione delle provenienze dei “proverbi non italiani” appartenenti alle concordanze di *proverbio* da itWaC.

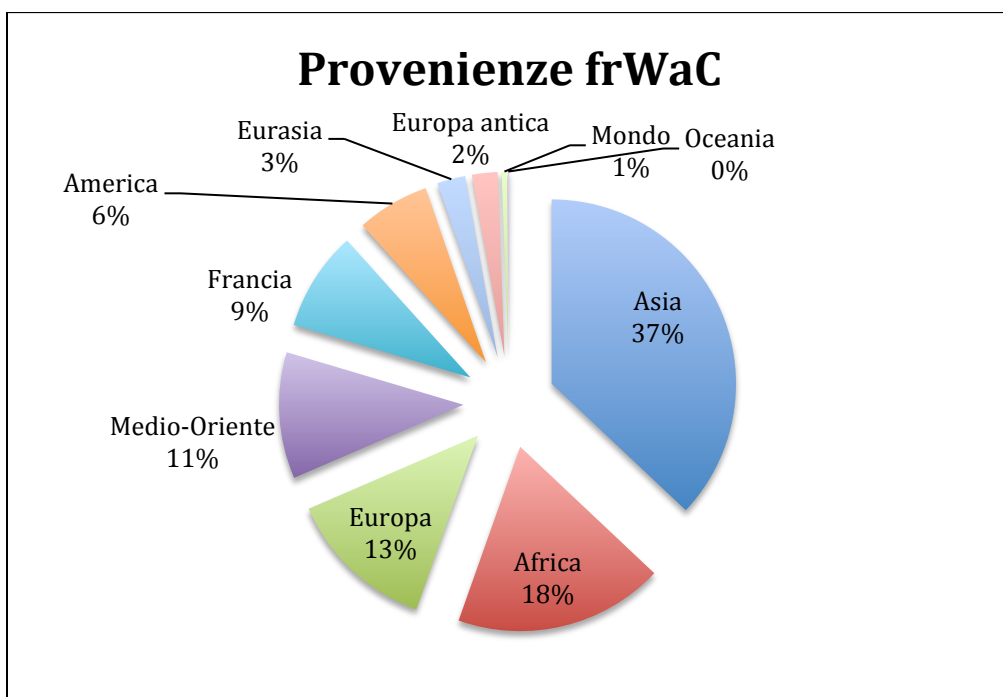


Fig. 7. Distribuzione delle provenienze dei “proverbi non francesi” appartenenti alle concordanze di *proverbe* da frWaC.

I dati parlano da soli: l’Asia, con a capo la Cina, è un riferimento extraculturale comune a entrambi i popoli. Questo spiega la ragione per cui nelle due tabelle di Fig. 4 e Fig. 5 ci sia come prima provenienza la Cina e nei grafici di Fig. 6 e Fig. 7 sia sempre l’Asia ad

occupare il primo posto per percentuale di occorrenze. L'Asia viene dunque considerata nella sua veste di generatrice di saggezza popolare e filosofica, il che le impone di essere citata attraverso i suoi proverbi, che non sono solamente cinesi, ma anche indiani, giapponesi, vietnamiti e tibetani, tra gli altri.

Dopo tale riferimento comune, si entra nelle specificità di ognuno dei due popoli analizzati: per gli italiani il secondo riferimento culturale è dato dai proverbi provenienti dai paesi dell'Europa, per i francesi, invece, si guarda all'Africa, un continente certamente lontano, ma anche alquanto vicino per ragioni storiche e politiche. In seguito, gli italiani volgono il proprio sguardo alle loro radici storiche e folkloriche o comunque alla tipicità delle zone d'origine, il che fa sì che i proverbi regionali e locali occupino il terzo posto, prima ancora dei proverbi africani. Per i francesi, invece, dopo l'Africa si passa dall'Europa e dal Medio-Oriente per arrivare solo in seguito alle regioni francesi. Le altre provenienze, come l'America, l'Eurasia, ma anche l'Europa antica (che racchiude dagli antichi greci e latini, finanche ai galli), l'Oceania e quell'insieme di provenienze presenti solo nel corpus francese che, data l'impossibilità di identificarle con una regione ben precisa, abbiamo chiamato "Mondo", sono simili in termini percentuali.

Cerchiamo di capire, in termini assoluti, per quali provenienze si differenziano i due corpora. In particolare, sarebbe interessante capire perché i francesi scelgono di enunciare più proverbi stranieri degli italiani e se tale differenza si concentra soprattutto in certe aree del mondo.

Si consideri la seguente tabella:

Provenienze	itWaC		frWaC	
Asia	218	34%	314	37%
Europa	113	18%	111	13%
Italia / Francia	95	15%	74	9%
Africa	73	11%	156	18%
Medio-Oriente	66	10%	94	11%
America	24	4%	54	6%
Eurasia	24	4%	21	2%
Europa antica	22	3%	19	2%
Oceania	2	0%	1	0%
Mondo	0	0%	4	0%

Le tendenze numeriche sono molto indicative. Lasciando per ora da parte le percentuali, i valori assoluti sono molto interessanti, in quanto indicano dove vi è una diversa concentrazione di proverbi non italiani o francesi. In grassetto abbiamo evidenziato le differenze più salienti: esse riguardano una presenza nettamente maggiore di proverbi asiatici, africani, medio-orientali e americani per i francesi, contro una presenza molto più ampia di proverbi regionali per gli italiani. Se per l'Asia abbiamo già dato una nostra spiegazione, per le altre provenienze possiamo avanzare le seguenti ipotesi: l'Africa, il Medio-Oriente e l'America sono le zone in cui si è concentrato il colonialismo francese, che impone oggi alla Francia di guardare alle popolazioni sottomesse fino a un passato tutto sommato recente come una fonte di saggezza ancora oggi viva. Questo è un po' quanto accade per l'Italia ma nei confronti delle proprie tradizioni locali. La Francia sembrerebbe guardare meno alle proprie tradizioni locali, forse perché può attingere dalle tradizioni di altri popoli sparsi per il mondo o forse perché si percepisce come uno stato molto più unito e omogeneo dell'Italia. Il raffronto, comunque, deve essere fatto proprio su questo piano: da un lato, abbiamo ancora un'Italia in cui vivono tradizioni locali antiche e che si percepisce come uno stato ancora frammentato tanto che ha bisogno di sottolineare che una fetta considerevole dei propri proverbi non è composta da proverbi italiani bensì locali; dall'altro lato, abbiamo una Francia molto più unita e dalle tradizioni più amalgamate, che continua a prendere spunto dalla saggezza dei popoli una volta soggiogati a essa. Insomma, si nota facilmente quanto anche la semplice presenza di proverbi accompagnati da certi marcatori che indicano qualche aspetto particolare, come la loro provenienza, sia un indice di come evolva la sapienza di un popolo e di come si percepisca un popolo. Da tale percezione, infatti, deriva la scelta dei riferimenti culturali e sapienziali del popolo stesso.

3.2.3. “Interesse storico-folkloristico”

Similmente a quanto abbiamo notato per i proverbi locali o dialettali, ci sono altri casi in cui i proverbi citati erano italiani ma venivano enunciati solamente per un mero interesse paremiologico oppure utilizzati come testimonianze di certe usanze, costumi, abitudini rurali (come i tempi della semina e del raccolto) o comportamenti sociali nel passato. Dato che questi proverbi venivano enunciati per una finalità metaparemiologica o per il semplice gusto di presentare usanze storiche o elementi folklorici, abbiamo deciso di dare a questa classe di esempi l'etichetta “Interesse storico-folkloristico”. Vediamo, di seguito, alcuni esempi.

- (10) Approfondimento sulla festa. La tradizione dice che sul Laterano passavano le streghe che si dirigevano verso il noce di Benevento (...). Riguardo alle lumache si diceva “Per ogni corna mangiata una sfortuna scongiurata”. E davanti ai piatti fumanti si cancellavano i litigi. Il **proverbio**: “Chi compra l'aglio per S. Giovanni ha fortuna tutti gli anni”.⁵¹
- (11) Il Sole sorge alle 6.07 e tramonta alle 20.25 / La Luna (nuova) si leva alle 4.59 e tramonta alle 20.33 / Oggi accadde: 4 agosto 1944 i nazisti catturarono la piccola ebrea Anna Frank e la sua famiglia / Compleanni: auguri ad Antonio Maccanico / Il **proverbio** - L'arte non ha peggior nemico dell'ignoranza / Il nome di oggi - Giovanni / All'origine del nome troviamo l'ebraico “Yohanan” o “Yeohanan”, formato da Yoh o Yah, abbreviazione di Yahweh, Dio e hanan.⁵²
- (12) Un **proverbio** “Vale più un vecchio in un canto che un giovane in un campo”. / Una parola / scolta: guardia, sentinella. / Una curiosità / La Pizia, indovina di Delfi, masticava le foglie di alloro per entrare in trance e i consultatori che avevano ottenuto dall'oracolo.⁵³
- (13) Il **proverbio** ‘dei disegni son piene le mura’ indica la facilità con cui si fanno progetti senza sicuro fondamento. ‘Non bisogna far disegni sulla roba degli altri’, è infine un invito a non contare sui beni e sugli aiuti altrui.⁵⁴

I quattro esempi mostrati indicano le ragioni per cui si è eliminata la categoria di proverbi che abbiamo chiamato “Interesse storico-folkloristico”. Notiamo, infatti, che in (10) si racconta dapprima un’usanza del passato, ossia una festa legata al culto di San Giovanni, per poi presentare un proverbio che aggiunge solamente particolari legati allo stesso culto: non vi è, dunque, altro intento comunicativo se non quello didattico-esplicativo e folklorico. In (11), al contrario, si considera il proverbio come una semplice curiosità del giorno, come potrebbe essere, ad esempio, il santo del giorno, oppure gli orari dell’alba e del tramonto o ancora gli eventi storici accaduti lo stesso giorno nel passato. Si tratta di un uso dei proverbi molto simile a quello presente nell’esempio (12), in cui il proverbio viene mostrato come una

⁵¹ URL: <<http://blog.libero.it/coMizia/view.php?id=coMizia\&gg=0\&mm=0506>>. Link non più attivo. Grassetto nostro.

⁵² URL: <<http://www.tgcom.it/tgmagazine/articoli/articolo211744.shtml>>.

⁵³ URL: <<http://www.italian.it/giornaliero/28Apr98.html>>.

⁵⁴ URL: <<http://www.educational.rai.it/lemma/testi/pittura/diseño.htm>>.

mera curiosità. In (13), infine, gli autori del testo hanno scelto un campo semantico (quello del disegno), decidendo di mostrare proverbi, espressioni idiomatiche e altro ancora relativi allo stesso ambito semantico.

3.2.4. Proverbi appartenenti a opere letterarie anteriori al XX secolo

In seguito, abbiamo cominciato a notare la presenza di proverbi appartenenti a opere letterarie antiche o comunque anteriori al XX secolo. Spesso facevano riferimento al *Libro dei Proverbi* della Bibbia (capitava anche che la parola proverbio non fosse seguita da alcun proverbio ma semplicemente indicasse questo libro; come vedremo, questo caso è stato catalogato come un esempio di “no proverbio”) o semplicemente ad altri passi della Bibbia, ma anche a opere come *I promessi sposi* di Alessandro Manzoni o a *Le astuzie di Bertoldo e le semplicità di Bertoldino* di G. C. Croce, o ancora a *I Malavoglia* di Giovanni Verga, al *Don Chisciotte della Mancia* di Miguel de Cervantes, al *Canzoniere* di Francesco Petrarca, ad alcune commedie di Carlo Goldoni, alla *Mandragola* di Niccolò Machiavelli, al *Convivio* di Dante Alighieri, *inter alia*. Dato che il nostro obiettivo consisteva nello studio di proverbi italiani e francesi effettivamente in uso, abbiamo scelto di eliminare questi esempi perché appartenevano a opere non contemporanee e non garantivano, da soli, l’effettivo utilizzo degli stessi ai nostri giorni.

In relazione al corpus frWaC, abbiamo trovato diversi esempi di proverbi provenienti da opere antiche. Anche per il francese, tra tali opere sono presenti la Bibbia (l’Antico e il Nuovo Testamento ricoprono una parte importante di questi esempi), il *Don Chisciotte della Mancia* tradotto in francese, il *Discours sur Homère* di Houdard de La Motte, e qualche lettera di Voltaire. Si vedrà in § 3.3 che la presenza di proverbi tratti da opere letterarie anteriori al XX secolo è maggiore per i risultati di itWaC che per frWaC.

Crediamo sia opportuno aprire una parentesi metodologica su come abbiamo proceduto all’individuazione di questa categoria di esempi. È, infatti, opportuno evidenziare che anche questa fase è stata compiuta manualmente, in quanto non ci sono etichette formali che indichino se un proverbio sia desueto e utilizzato in opere più o meno antiche. Sebbene alcuni paremiologi segnalino che i proverbi si adeguano al passare del tempo, ammodernando il proprio lessico o la propria struttura semantica, in realtà ciò non è sempre valido: pertanto, non è detto che a strutture o a un lessico arcaici corrisponda un proverbio in disuso. Quindi, per riconoscere i contesti appartenenti alla Bibbia o a opere letterarie precedenti il XX secolo, abbiamo utilizzato due modi: ci siamo concentrati sugli URL (metodo più automatico) oppure

abbiamo letto tutto il contesto; in caso di dubbio, abbiamo cercato in Internet quei contesti che ci sembravano più desueti. Se il link fornito dal corpus non era più attivo, tali ricerche venivano effettuate con dei copia e incolla di porzioni del contesto in motori di ricerca.

3.2.5. “No proverbio”

Abbiamo ancora trovato, tanto tra le concordanze di itWaC quanto tra quelle di frWaC, esempi in cui la parola *proverbio* non era utilizzata in funzione di *marcatore di proverbio*: essa, dunque, non accompagnava l'enunciazione di alcun proverbio. Pertanto, siccome questi esempi non fornivano alcun proverbio, abbiamo deciso di eliminarli. Come si vedrà successivamente in § 3.3, questi casi ricoprono una percentuale abbastanza esigua considerando la quantità di proverbi tanto italiani e francesi quanto non italiani e non francesi ottenuta utilizzando tale *marcatore di proverbio*, cosa che indica che le parole *proverbio* e *proverbe* permettono effettivamente di ottenere una quantità considerevole di proverbi utilizzati nella comunicazione.

Per spiegare meglio quanto detto, proporremo, qui di seguito, alcuni degli esempi eliminati⁵⁵.

- (14) Mio padre compra un calendario nuovo su cui si scrive anche il gorno [giorno] lunare e solare. Ogni pagina del calendario c'e' [c'è] un mondo spirituale, si trova anche una poisia [poesia], un **proverbio**, le parole sagge, il segno.⁵⁶
- (15) di Giuliano Vignini 192 pag., Euro 10.50 - Paoline Editoriale Libri (Bibbia Paoline n. 8) (...) Le prime righe INTRODUZIONE A **PROVERBI** CARATTERISTICHE GENERALI / Nell'AT il termine masal (**proverbio**) assume sia una valenza simbolica (parabola, similitudine, allegoria, ecc.) sia un significato specifico (detto o sentenza popolare).⁵⁷
- (16) Dal **proverbio** al racconto / Per questa proposta si suggerisce ai ragazzi di scegliere un proverbio e di svilupparlo in senso letterale, a prescindere perciò dalla sua interpretazione in senso traslato.⁵⁸

⁵⁵ Gli esempi sono stati copiati e incollati tali quali sono stati reperiti nel corpus itWaC. Eventuali correzioni sono state aggiunte in parentesi quadre. Le uniche aggiunte sono state fatte tra parentesi quadre per rendere più chiari gli esempi stessi. Il nodo, rappresentato dalla parola *proverbio*, è stato evidenziato in grassetto. Gli “a capo” sono stati segnalati con il segno “/”.

⁵⁶ URL: <<http://www.turistipercaso.it/viaggi/itinerari/testo.asp?ID=1806>>.

⁵⁷ URL: <<http://www.cafeletterario.it/267/883152402X.htm>>.

⁵⁸ URL: <<http://kidslink.scuole.bo.it/irrsaer/prosus/miniatur1.html>>.

- (17) Quando arriva la mezza età, è la moglie che se ne va (14/06/2004) / Non è un nuovo **proverbio**, ma il risultato di uno studio condotto negli Stati Uniti sul divorzio dei baby-boomers.⁵⁹
- (18) Così i protagonisti del video e il frenetico inutile agitarsi diventano un simbolo in movimento, efficace come un **proverbio** ed immediato come un logotipo, del paradosso dell'esistenza.⁶⁰
- (19) Il caro, vecchio Giovanni [Trapattoni] ha anche inventato un linguaggio: un idioma che mescola una sintassi colorata al **proverbio** orecchiato, la frase che, soltanto apparentemente, non possiede significato.⁶¹
- (20) Lui diceva che nei **proverbi** c'era qualcosa di triste e decadente. I **proverbi** sono quelli da sempre, o per lo meno da molti decenni e questo voleva dire che l'uomo non cambiava mai, rimaneva sempre lo stesso, tale e quale. Il **proverbio** dice all'uomo che gli è preclusa ogni possibilità di cambiamento. Mentre le barzellette cambiano di generazione in generazione, e quelle che facevano ridere venti anni fa magari oggi non fanno più ridere e ne fanno ridere invece altre che anni addietro.⁶²
- (21) Spesso la firma in un messaggio, oltre a riportare il nostro nome e cognome, può contenere anche una frase, un detto, un **proverbio**... Insomma, qualcosa che in qualche modo ci caratterizzi.⁶³

Notiamo che in nessuno degli esempi appena riportati è presente alcun proverbio. In (14) la parola proverbio fa riferimento semplicemente a un servizio offerto da un calendario, alla stregua di poesie, espressioni sapienziali, delle fasi solari e lunari. In (15), invece, non ci si riferisce ai proverbi *tout court*, bensì al *Libro dei Proverbi* della Bibbia. In (16), si espone un esercizio didattico relativo ai proverbi senza presentare alcun proverbio, mentre in (17) si fa riferimento a una frase esplicitando che non si tratta di un proverbio. L'esempio (18) riporta una similitudine tra un simbolo di un qualcosa che non conosciamo e l'efficacia del proverbio. (19) si riferisce al linguaggio colorito dell'allenatore di calcio Giovanni Trapattoni

⁵⁹ URL: <http://miaeconomia.virgilio.it/retrieval/home/archivio_articoli.aspx?idchannel=21&idcategory=163&pagenum=5&can=SOLDI+E+FAMIGLIA&cat=Matrimonio+e+patrimonio>.

⁶⁰ URL: <http://www.genovagando.it/eventi/approfondimenti/Old/Sergio_Muratore_da_Joyce%20&%20Co.htm>.

⁶¹ URL: <<http://www.liberalfondazione.it/archivio/fl/numero2/rap.htm>>.

⁶² URL: <<http://www.pennadautore.it/alboro/racconti/2004/daniele.htm>>.

⁶³ URL: <http://www.uiciechi.it/servizi/riviste/TestoRiv.asp?id_art=1698>.

che farcisce i propri discorsi di proverbi, mentre (20) propone una serie di considerazioni attorno al proverbio in generale senza enunciarne nemmeno uno. Infine, in (21) si spiega la possibilità di aggiungere automaticamente in calce a una e-mail, oltre alla firma dello scrivente, anche una frase, un detto o un proverbio che possa caratterizzare l'autore della missiva.

3.2.6. “Ripetizioni”

L'altra categoria di risultati scartati è composta dalle ripetizioni di esempi. Si è affermato nel precedente capitolo che, sebbene nella costruzione dei corpora itWaC e frWaC si sia riposta una grande attenzione nell'eliminazione di duplicati di testi, alcuni di questi duplicati sono impossibili da individuare e scartare automaticamente. Si tratta, frequentemente, di stringhe di testo che rimangono uguali in testi diversi (è il caso dell'oggetto di discussioni di forum che rimane invariato per tutte le risposte dei diversi utenti: se il proverbio si trova nell'oggetto, questo viene ripetuto per il numero di risposte presenti nel forum stesso), oppure di testi uguali presenti in siti diversi aventi URL differenti (è il caso di articoli copiati e incollati in siti diversi) o ancora testi quasi uguali che si distinguono soltanto per un link oppure per un'immagine (differenze, queste, che non influiscono sul testo stesso). L'individuazione di duplicati di testi viene facilitata ordinando gli esempi secondo il contesto a destra del nodo: questo ordine, infatti, aiuta l'occhio umano a individuare quelle concordanze esattamente uguali. Ciononostante, sebbene abbiamo posto la massima cura nell'eliminazione di queste ripetizioni, è possibile che qualcuna di queste sia passata inosservata.

Tutti questi casi di ripetizioni sono stati eliminati. Le ragioni di questa scelta non stanno, come per i casi sopramenzionati, nella non adeguatezza dei proverbi individuati, bensì nel disturbo che potrebbero arrecare alla successiva fase di analisi quantitativa dei risultati ottenuti: se, infatti, un proverbio è ripetuto un certo numero di volte, ma si nota che il contesto che segue e precede il proverbio stesso è sempre il medesimo, ciò non significa che esso sia più frequente di un altro proverbio che ha meno occorrenze ma che ricorre in contesti sempre diversi e che viene utilizzato da utenti diversi. Dato che uno dei nostri obiettivi consisteva nel proporre una lista di frequenza dei proverbi, queste ripetizioni avrebbero falsato le nostre future considerazioni statistiche, pertanto si è proceduto alla loro eliminazione già a partire da questa fase.

3.3. Analisi quantitativa delle concordanze validate e di quelle eliminate

Abbiamo passato al vaglia le concordanze di entrambi i corpora, per suddividere gli esempi di un certo interesse da quelli meno interessanti. Tale operazione è stata effettuata secondo le regole precedentemente delineate, applicate, così come le abbiamo esposte, per realizzare la scrematura delle concordanze di entrambi i corpora. Poniamo l'accento sul carattere semiautomatico dell'operazione: ciò significa che è stata necessaria un'attenta lettura dei contesti d'uso per ogni concordanza reperita.

Sarebbe interessante studiare la distribuzione quantitativa delle diverse tipologie di concordanze individuate: anzitutto, perché si tratta di una maniera per confrontare i due corpora al fine di evidenziare eventuali somiglianze e differenze e, poi, perché valori simili tra i due corpora possono essere un'ulteriore prova dell'elevato grado di compatibilità e comparabilità degli stessi e dell'uso del marcatore di proverbio individuato per entrambe le lingue.

Ci concentreremo, a questo punto, su alcuni dati che rappresentano la distribuzione delle diverse categorie di proverbi individuate: andremo dal nostro obiettivo, ovvero i proverbi italiani effettivamente in uso oggi giorno, per poi presentare le percentuali dei risultati eliminati. Prima di passare all'esposizione di tali dati, è opportuno sottolineare che il lavoro di suddivisione delle diverse concordanze secondo i criteri suesposti è stato tutt'altro che semplice, breve o privo di esempi dubbiosi che erano al confine tra le diverse categorie individuate. Non di rado ci siamo imbattuti in proverbi non italiani o non francesi ripetuti e magari appartenenti a opere letterarie precedenti il 900 (pertanto al confine tra tre categorie: "proverbi non italiani", proverbi "ripetuti" e proverbi di "Letteratura precedente al 900"). Un esempio può essere portato da alcuni proverbi tratti dal *Don Chisciotte* e reperiti tanto in itWaC quanto in frWaC. Come assegnare una sola categoria a proverbi che in realtà sarebbero appartenuti a diverse categorie? Abbiamo deciso di istituire una sorta di "precedenze": in presenza di più categorie allo stesso tempo, la categoria "ripetizioni" sarebbe stata quella sovraordinata, seguita in ordine dalla categoria "opere letterarie anteriori al 900" e da quella composta da proverbi non italiani o non francesi. Lo stesso discorso vale per quando eravamo in presenza contemporaneamente di categorie come "ripetizioni" e "no proverbio": abbiamo sempre dato la precedenza alla categoria ripetizioni. Ciononostante, molti di questi esempi potevano avere diverse interpretazioni e, come si sa, in ambito linguistico non sempre esiste il bianco e il nero: le sfumature ricoprono aree molto vaste e per questo le imperfezioni devono

essere accettate. Inoltre, non potendo eseguire un'analisi di questo tipo in maniera automatica o "matematica", assumiamo che ci sia un piccolo grado di approssimazione.

Vediamo ora come abbiamo ricavato le percentuali riguardanti le diverse occorrenze delle parole *proverbio* e *proverbe*. Anzitutto, abbiamo definito la percentuale dei due gruppi più numerosi e al contempo più facilmente calcolabili: i proverbi che abbiamo mantenuto per la nostra analisi, categoria a cui daremo l'etichetta di "proverbi adeguati all'analisi", e i proverbi stranieri, che in realtà comprendono anche i proverbi regionali, dialettali o locali, che abbiamo definito con le etichette di "proverbi non italiani" e "proverbi non francesi". La quantità dei "proverbi adeguati all'analisi" emerge chiaramente dal semplice computo dei risultati ottenuti dopo la fase di scrematura dei due corpora: essi sono 1.110 per il corpus italiano, che equivale al 29,9% delle concordanze ottenute, e 974 per quello francese, pari al 27,6% delle concordanze della parola *proverbe*. Con qualche arrotondamento, possiamo affermare che i proverbi italiani che analizzeremo, reperiti in questa prima fase di ricerca, sono il 30% delle concordanze ottenute, mentre quelli francesi equivalgono al 28%.

Se avevamo un numero maggiore di risultati utili alla nostra analisi provenienti da itWaC rispetto agli stessi riguardanti frWaC, cosa che probabilmente è dovuta alla quantità maggiore di concordanze ottenute con itWaC (3.706) rispetto a quelle ottenute con frWaC (3.537), la situazione si ribalta considerando i "proverbi non italiani / francesi". I cosiddetti "proverbi non italiani", infatti, sono 637, pari al 17% delle concordanze totali della parola *proverbio*, mentre i "proverbi non francesi" sono 848, pari al 24% delle concordanze della parola *proverbe*. Possiamo, pertanto, ipotizzare che in francese l'enunciazione della parola *proverbe* farà emergere una quantità di proverbi francesi di poco superiore rispetto a quelli non francesi (28% dei proverbi francesi contro il 24% di quelli non francesi), mentre in italiano il divario tra proverbi italiani e proverbi non italiani è più pronunciato: c'è uno scarto di 13 punti percentuali tra queste due categorie, avendo una percentuale del 30% per i proverbi italiani e una percentuale del 17% per quelli non italiani. Probabilmente ciò significa che, in percentuale, i parlanti francesi utilizzano più proverbi stranieri rispetto agli italiani. Notiamo, infatti, che le percentuali dei proverbi francesi e italiani sono tutto sommato simili (si tratta del 28% contro il 30%), il che dimostra un'elevata somiglianza comparabilità dei due corpora, ma anche dell'uso del marcatore di proverbi individuato per la lingua italiana e per quella francese. Per questa ragione, lo scarto di 17 punti percentuali relativo all'uso dei proverbi non italiani o non francesi a favore dei risultati ottenuti con frWaC altro non fa che

sottolineare un uso maggiore di “proverbi non francesi” da parte dei parlanti francesi rispetto all’uso dei “proverbi non italiani” da parte dei locutori italiani. Queste considerazioni, ovviamente, valgono per l’enunciazione di proverbi concomitante all’enunciazione del marcatore di proverbi individuato per le due lingue in oggetto.

A questo punto, è possibile determinare la fetta di concordanze delle parole *proverbio* e *proverbe* composta da proverbi “veri”, tanto italiani o francesi, quanto regionali o stranieri: essa ricopre rispettivamente il 47% e il 52% dei corpora itWaC e frWaC. Pertanto, in assoluto, si utilizzano più proverbi in francese che in italiano. Questo scarto è senz’altro da imputare alla maggiore presenza di proverbi non francesi rispetto a quella dei proverbi non italiani.

In seguito, abbiamo definito le percentuali delle ulteriori categorie che ricoprono, insieme, il 53% delle concordanze di itWaC e il 48% di quelle di frWaC. Si riassumono nella seguente tabella le percentuali relative a ogni categoria in cui abbiamo suddiviso le concordanze di itWaC e frWaC.

Tipologia concordanza	itWaC concordanza: <i>proverbio</i>	frWaC concordanza: <i>proverbe</i>
Proverbi adeguati all’analisi	30%	28%
Proverbi non italiani / non francesi	17%	24%
Interesse storico-folkloristico	9%	8%
Letteratura precedente al 900	11%	2%
No proverbio	15%	14%
Ripetizioni	18%	24%

Fig. 8. Tabella che riassume la distribuzione delle diverse categorie in cui abbiamo suddiviso le concordanze delle parole *proverbio* e *proverbe*. In grassetto si sono evidenziati i dati più diversi.

Notiamo, a questo punto, ulteriori somiglianze e disuguaglianze: se in francese si usano più “proverbi non francesi”, in italiano si utilizzano parecchi proverbi appartenenti a opere letterarie antecedenti al XX secolo: questo probabilmente è dovuto alla forte presenza di opere antiche presenti anche in versione integrale nei siti Web italiani. Le categorie relative all’interesse storico folkloristico, invece, sono praticamente uguali, pertanto si può sostenere che tanto gli italiani quanto i francesi utilizzino proverbi per parlare del passato o delle tradizioni più folkloristiche nella stessa misura. Si utilizza anche la parola *proverbio* senza che

sia accompagnata dall'enunciazione di uno o più proverbi secondo le stesse percentuali. Un discorso a parte deve essere fatto per le ripetizioni, le quali sono più presenti in francese che in italiano ma che riguardano la composizione dei due corpora piuttosto che l'utilizzo dei proverbi assieme a marcatori di proverbi.

3.4. Natura dei proverbi individuati

Prima di passare alle fasi successive, è opportuno soffermarsi sulla natura dei risultati ottenuti. Si è detto, all'inizio di § 3.2, che l'unica cosa che hanno in comune i proverbi individuati è il fatto di essere definiti proverbi da parte degli autori dei testi in cui essi compaiono. Non dobbiamo dare per scontato che l'utente medio della Rete (tale considerazione può essere estesa al parlante medio) sappia che cosa sia un proverbio, dal momento che, come ribadito nel primo capitolo⁶⁴, nemmeno i linguisti sono unanimemente concordi sulla definizione di proverbio. Rispetto a quanto emerso fino ad ora in letteratura, il presente lavoro può aggiungere alcune considerazioni importanti sulla competenza paremiologica del parlante medio. Anche se si approfondirà questo punto successivamente, desideriamo accennare che ciò che i parlanti definiscono proverbio può essere in realtà un aforisma di qualche personaggio del passato o contemporaneo, oppure una semplice espressione idiomatica. Posto che uno dei nostri obiettivi consiste nell'analizzare i proverbi italiani e francesi effettivamente in uso oggi giorno per finalità legate non solo alla paremiologia, ma anche al mondo della traduzione e dell'interpretazione, abbiamo deciso di considerare tali risultati veri e propri proverbi. Se pensiamo al lavoro del traduttore e dell'interprete, capiamo che entrambi hanno a che fare con la lingua in uso e non con la lingua in teoria. Se vogliamo dare risposte concrete che possano aiutare queste figure professionali nel loro lavoro, dobbiamo considerare ciò che accade nella realtà, senza trascurare ciò che dovrebbe accadere in teoria, così da fornire dati tanto pratici quanto teorico-normativi.

Questo punto ha sollevato non poche obiezioni nel corso del 9th *Interdisciplinary Colloquium on Proverbs* del 2015. La nostra proposta consistente nell'analizzare anche quegli elementi ritenuti proverbi dai parlanti ma, di fatto, non rispondenti alle caratteristiche del proverbio, ha diviso il pubblico tra favorevoli e contrari: per i sostenitori, la cui principale portavoce è stata la paremiologa finlandese Liisa Granbom-Herranen, è importante tenere in considerazione e studiare che cosa i parlanti ritengano che sia il proverbio; per gli oppositori,

⁶⁴ Cfr. § 1.1.

si tratta invece di quanto di più sbagliato possa esistere in paremiologia, in quanto il paremiologo non può abbandonare le proprie convinzioni a favore di errate considerazioni comuni. Anche in questo caso, ci troviamo di fronte a due approcci notevolmente diversi, che possono essere entrambi validi a seconda dei punti di vista ma che non possono essere esclusi a priori sulla base di preconcetti. Probabilmente, dovremmo redimere la questione rifacendoci a un proverbio latino: *in medio stat virtus*.

Per essere coerenti con le nostre scelte e con le norme di cui ci siamo dotati, abbiamo fatto decidere al marcatore di proverbio e al contesto gli esempi da conservare e quelli da eliminare. Pertanto, se un esempio che conteneva una frase o un'espressione che non potevano essere considerati proverbio soddisfaceva tutte le caratteristiche contestuali da noi individuate, abbiamo deciso di mantenerlo alla stregua di qualsiasi "vero" proverbio.

Abbiamo scelto tale approccio anche per un'altra motivazione: prendere in esame tutti i proverbi che sono ritenuti tali dagli enunciatori è l'unico metodo per capire in che direzione si stia muovendo la paremiologia, segnatamente nel caso di invenzione di nuovi proverbi oppure di modifica intenzionale o involontaria di proverbi già esistenti. L'analisi di dati reali permette anzitutto di capire che i proverbi sono attualmente utilizzati in diverse varietà diastratiche sia in italiano sia in francese, il che contrasta con le opinioni diffuse secondo cui il proverbio stia scomparendo, non venga più utilizzato e che non si creino più nuovi proverbi.

Ad ogni modo, è indubbio che molti dei proverbi ottenuti non siano in realtà proverbi oppure che altre frasi che avrebbero tutte le caratteristiche dei proverbi non siano attestate nei più comuni dizionari di proverbi. In quest'ultimo caso, si potrebbe pensare che effettivamente essi non siano proverbi, oppure che essi siano stati creati di recente o ancora che siano davvero proverbi ma che non siano ancora entrati nei dizionari. Questo impone una seria riflessione, da un lato, sulla competenza paremiologica dei parlanti, e, dall'altro, sulla natura dei dizionari di proverbi esistenti.⁶⁵

Qualunque sia la ragione di questi fatti, è certo che la sola formulazione di simili interrogativi è stata possibile grazie all'approccio *corpus-driven* da noi scelto che ci ha fatto considerare tutti i risultati validati grazie all'analisi del contesto precedente e successivo alla parola *proverbio* e *proverbe*. Ribadiamo che tale approccio è il solo in grado di evidenziare l'andamento reale della lingua anche in materia di paremiologia.

3.5. Scelta delle informazioni da mantenere per ogni proverbio reperito

⁶⁵ Cfr. a tal proposito § 4.5 e 4.7.1.

Dopo aver capito come scremare le concordanze italiane e francesi, e prima di passare al vaglio i due corpora, abbiamo stabilito quali fossero i dati che avremmo voluto evidenziare per ognuno dei risultati utili per la nostra analisi.

Per prima cosa, ci sembrava opportuno mantenere gli URL di ogni esempio, sebbene avessimo notato da subito che molti di questi non fossero più attivi, sebbene la quantità di URL inattivi sia maggiore per itWaC. Probabilmente, la ragione di ciò è dovuta al fatto che i testi presenti in frWaC sono leggermente più recenti rispetto a quelli presenti in itWaC.

Per giungere a considerazioni legate al periodo medio di pubblicazione dei testi in Internet, abbiamo dovuto evidenziare questo tipo di informazione. Per ogni risultato, abbiamo quindi indicato la data del testo in cui appariva, non solo per capire la ragione per cui molti link non erano più validi (cosa che, in fondo, riveste uno scarso interesse ai fini della nostra ricerca), ma anche per essere sicuri di analizzare testi prodotti negli ultimi anni. Questa informazione avrebbe avuto una duplice rilevanza: da un lato, sarebbe stata la prova del periodo di attestazione dei nostri proverbi (utile anche per ulteriori studi diacronici oppure per testare l'evoluzione dei proverbi), dall'altro avrebbe evidenziato la quantità di proverbi che ancora oggi vengono impiegati, nonché la loro frequenza d'uso. Tuttavia, non sempre l'informazione riguardante la datazione era disponibile. Nel caso in cui non lo fosse, abbiamo scelto di indicare la mancanza dell'informazione con un punto interrogativo. Per reperire questo dato, ci siamo basati sui contenuti presenti nei diversi contesti di 250 caratteri, oppure sugli URL (in cui spesso è indicata la data o almeno l'anno, un altro motivo dunque per ritenere di grande interesse l'URL), o ancora abbiamo optato, qualora fosse possibile, per ritrovare la pagina Web corrispondente in cerca di una qualche datazione.

Vediamo, dunque, qual è la datazione media dei due corpora. Se consideriamo il corpus itWaC, escludendo i casi in cui non siamo riusciti a trovare una data certa da attribuire ai testi (questi casi rappresentano la maggioranza con 860 occorrenze), notiamo che i testi si distribuiscono soprattutto tra il 2005 (con 92 occorrenze) e il 2002 (con 25 occorrenze). In relazione al corpus frWaC, abbiamo contato 674 casi in cui non siamo riusciti a reperire la datazione utilizzando i metodi appena esposti (ovvero considerando il contesto, l'URL oppure aprendo le pagine Web corrispondenti). Escludendo questi casi, notiamo che la maggior parte dei testi si concentra tra il 2007 (con 150 occorrenze) e il 2005 (con 28 occorrenze). Si vedano i dati in Fig. 9 (datazioni di itWaC) e in Fig. 10 (datazioni di frWaC). I dati sono ordinati per numero di occorrenze relative a ogni anno da noi reperito (per chiarezza, nelle due tabelle

abbiamo scelto di indicare solamente l'anno, sebbene avessimo indicato nella nostra banca dati, qualora fosse possibile, non solo l'anno di pubblicazione dei testi, ma anche il mese e il giorno):

itWaC	
DATAZIONE	OCCORRENZE
?	860
2005	92
2004	61
2003	39
2002	25
2001	16
2000	7
2007	2
1998	2
1999	1
1997	1
1991	1
1962	1
1937	1
1921	1

Fig. 9. Datazione dei proverbi reperiti in itWaC tramite la concordanza della parola *proverbio*.

frWaC	
DATAZIONE	OCCORRENZE
?	674
2007	151
2006	62
2008	45
2005	28
2004	6
2002	4
2003	1
2000	1
1999	1
1986	1

Fig. 10. Datazione dei proverbi reperiti in frWaC tramite la concordanza della parola *proverbe*.

Oltre a queste due informazioni, abbiamo creduto opportuno indicare anche la tipologia del testo in cui trovavamo i proverbi. Le principali categorie in cui abbiamo suddiviso i nostri testi sono state: “Articolo”, “Forum”, “Blog”. Accanto a queste categorie di

testi abbiamo altresì individuato altre tipologie di testi, come “Discorso” e “Letteratura”. Vediamo nel dettaglio a cosa corrisponde questa classificazione. Non pretendiamo di arrivare a una classificazione perfetta o indiscutibile, specialmente perché con un corpus così vasto le tipologie testuali sono davvero molto numerose. Pertanto, le nostre etichette non devono essere prese alla lettera, ma devono essere considerate grandi categorie generali secondo i principi che verranno ora esposti.

In primo luogo, sotto l’etichetta “Articolo” abbiamo incluso tutti quei testi monologici scritti, per lo più pubblicati in giornali o riviste di qualsiasi tipologia (dalle testate più conosciute a riviste meno note) e che non appartenessero a blog o forum.

In secondo luogo, sotto la voce “Blog” abbiamo categorizzato testi monologici spesso scritti in prima persona, o comunque con frequenti riferimenti in prima persona all’autore del testo o alle sue opinioni, con uno stile piuttosto colloquiale (sebbene non sia questa una *conditio sine qua non*) e determinati nella quasi totalità dei casi dalla presenza della parola *blog* all’interno dell’URL o comunque nel contesto.

Successivamente, è stata assegnata l’etichetta “Forum” a quei testi aventi quasi sempre una lunghezza totale più ridotta rispetto ai precedenti, in cui era spesso visibile l’oggetto della discussione oltre alla data e all’orario di caricamento del messaggio in Internet. Sono testi monologici benché alquanto prossimi al confine con i testi dialogici, in quanto si tratta, non di rado, di risposte a interventi precedenti o a un tema di discussione comune. Inoltre, è spesso indicata anche nell’URL la parola *forum*.

Le ulteriori categorie che abbiamo individuato sono, rispettivamente, quella denominata “Discorso” e quella etichettata come “Letteratura”. Sono categorie marginali rispetto alle tre precedentemente esposte ma comunque presenti e per questo individuate con due nomi distinti.

La categoria “Discorso” fa riferimento agli unici testi scritti ma più vicini al parlato monologico in ambito istituzionale. Si tratta di diverse tipologie di discorsi (soprattutto politici, ma non solo) pronunciati davanti a un pubblico a cui si fa riferimento nei testi stessi. Teniamo a sottolineare che non si tratta di trascrizioni, per altro mai presenti nei due corpora utilizzati nell’ambito del presente lavoro, bensì di discorsi scritti e utilizzati per essere pronunciati davanti a un pubblico.

L’ultima categoria individuata, definita “Letteratura”, fa riferimento a stralci di opere letterarie, teatrali, di finzione, tra le altre, scritte a partire dal XX secolo fino ai giorni nostri.

Contrariamente alle altre opere letterarie precedenti al XX secolo, abbiamo creduto opportuno mantenere questi testi in quanto comunque rispecchiano un utilizzo comunicativo del proverbio ai nostri giorni, sebbene si tratti di un utilizzo dei proverbi più artificioso rispetto a quelli che possiamo trovare in un articolo, oppure in blog o, ancora, in forum. Resta inteso che anche se questi proverbi appartenevano alla categoria “Letteratura”, non è detto che non potessero essere rimossi: se mostravano caratteristiche in comune con i proverbi eliminati (se erano, ad esempio, proverbi appartenenti alla categoria “Interesse storico-folkloristico” oppure proverbi non italiani), essi venivano a loro volta cancellati. Abbiamo così l’occasione per ribadire che l’eliminazione dei proverbi non è stata data dalle caratteristiche dei testi in cui i proverbi stessi si trovavano, bensì dalla tipologia del contesto in cui essi erano impiegati. L’analisi delle tipologie testuali in cui apparivano i proverbi è stata una fase successiva alla scrematura dei corpora. Ciononostante, per maggiore chiarezza, presentiamo già qui le nostre considerazioni.

Notiamo in entrambi i corpora una maggioranza di testi catalogati come articoli. Questo perché, effettivamente, tale tipologia testuale era molto presente negli esempi reperiti grazie alla concordanza delle parole *proverbio* e *proverbe* ma anche perché spesso non vi erano elementi sufficienti per attribuire con certezza ai vari contesti l’appartenenza a una tipologia o a un’altra. Ciò che emerge chiaramente da questa distribuzione di tipologie testuali è che la modalità di testo monologica è sicuramente quella che è presente in maggioranza. Ora, che si tratti di veri articoli di testate giornalistiche, siti internet oppure blog poco importa: ciò che conta ai fini dell’analisi degli esempi ottenuti è che la stragrande maggioranza dei proverbi reperiti, oltre a essere effettivamente stata reperita in testi scritti, appartiene a testi in cui un autore parla a lettori che vedranno successivamente il testo. L’effettiva minoranza di testi dialogici in tempo più o meno reale come i testi che si trovano nei forum sembrerebbe fare intendere che l’enunciazione dei proverbi è spontanea, ma anche che, almeno quella legata alla concomitante enunciazione di marcatori di proverbi, richiede, nella maggioranza dei casi, una certa pianificazione testuale. Queste osservazioni valgono tanto per i proverbi di itWaC quanto per quelli di frWaC: effettivamente, sebbene vi sia una notevole differenza riguardante i blog, molto più numerosi in frWaC che in itWaC, si tratta pur sempre di testi monologici, sicuramente più colloquiali, ma molto più vicini alla tipologia testuale dell’articolo.

Tipologia testuale	itWaC	frWaC
Articolo	787	612
Blog	87	225
Forum	172	105
Letteratura	56	30
Discorso	8	2

3.6. Importazione dati e costruzione della tabella di lavoro in Excel

Abbiamo già spiegato come effettuare la scelta delle concordanze da mantenere e quali dati registrare nella nuova banca dati di proverbi reperiti su itWaC e frWaC. Passiamo ora a illustrare come nella pratica si è proceduto alla costruzione di tale *database*.

Abbiamo deciso di non lavorare più online con il programma di gestione dei corpora WaCky, NoSketch Engine, utilizzando la funzione di salvataggio dati proposta dallo stesso programma al fine di salvare i dati in formato .txt. Abbiamo preso questa decisione in quanto la fase di pulitura dei due corpora sarebbe durata alcune settimane, il che imponeva di essere certi che il materiale con cui dovevamo lavorare (ossia i corpora stessi) non cambiasse in corso d'opera, nonché di avere sempre una connessione Internet attiva ed efficiente. In teoria, trattandosi di corpora ormai definiti, le concordanze di *proverbio* e *proverbe* non sarebbero dovute variare. Nondimeno, si sarebbe potuto inavvertitamente variare l'ordine di visualizzazione delle concordanze o la quantità di contesto da visualizzare. Inoltre, per avere la possibilità di lavorare anche offline, in maniera da ampliare il tempo da dedicare a questa fase, avremmo dovuto trovare un altro sistema. Per queste ragioni, abbiamo deciso di procedere al salvataggio in file di testo (l'unica modalità di salvataggio possibile assieme all'XML) delle concordanze di *proverbio* e *proverbe* per lavorare direttamente sui file salvati. Vediamo come si è proceduto per giungere a questi due file.

Per prima cosa, abbiamo cliccato sul comando "Save" nella colonna delle impostazioni a destra della schermata delle concordanze (cfr. Fig. 2). Si è così aperta una finestra con alcune opzioni di salvataggio (Fig. 11).

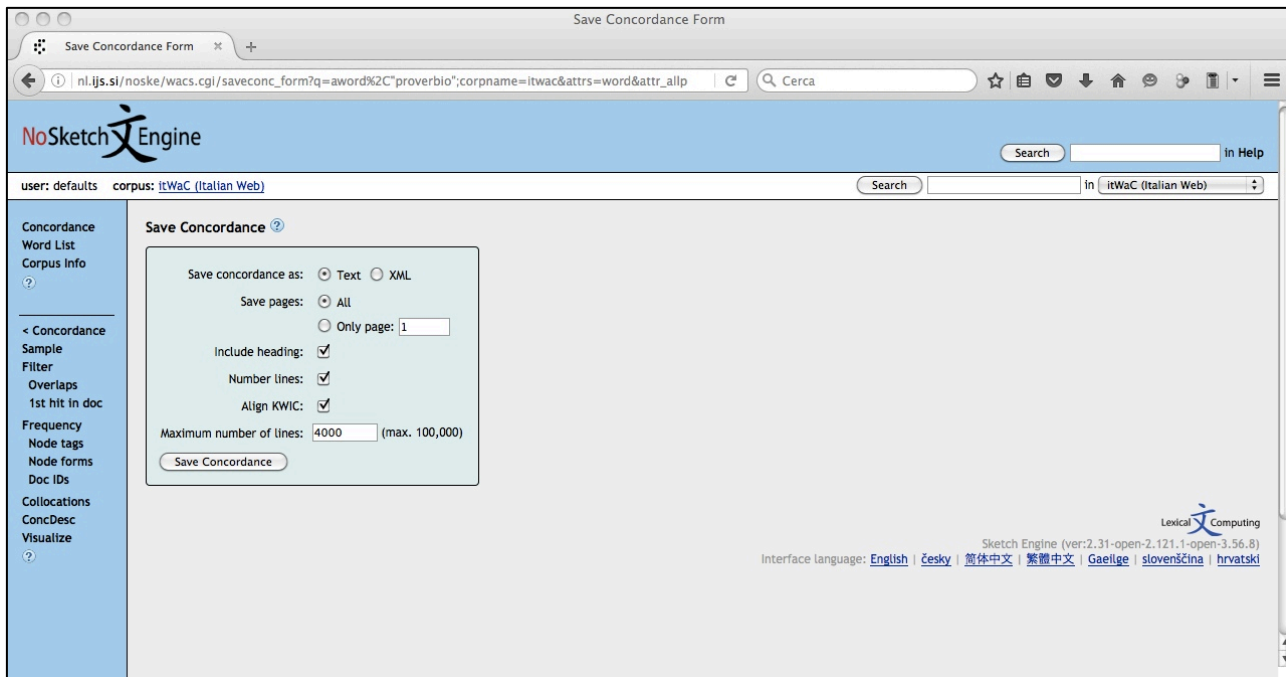


Fig. 11. Impostazioni di salvataggio delle concordanze.

Come si vede in Fig. 11, abbiamo scelto che il file su cui lavorare fosse un file di testo, che includesse l'intestazione (ovvero, l'URL) e il numero della riga, e che il cosiddetto KWIC (acronimo inglese di *keyword-in-context*), che corrisponde al nodo (nel nostro caso alla parola *proverbio*), fosse allineato. In seguito, abbiamo impostato come numero massimo di righe 4.000, sicuri che non avremmo mai raggiunto tale tetto, in quanto le occorrenze della parola *proverbio* nel corpus itWaC erano 3.706 e quelle di frWaC 3.537. Dopo aver scelto queste impostazioni, abbiamo cliccato su "Save Concordance" e scaricato il relativo file di testo visibile in Fig. 12.

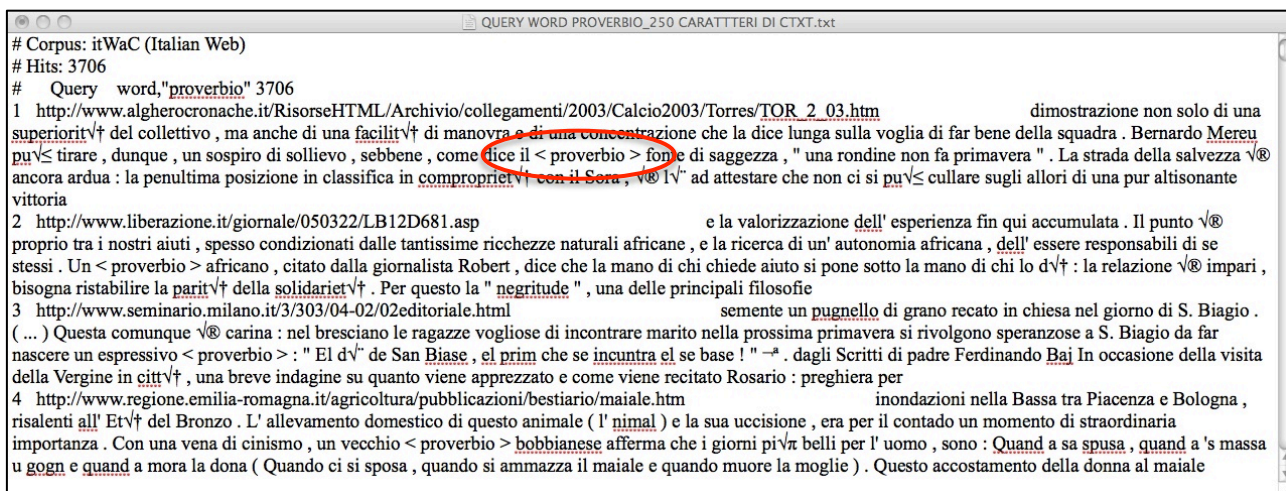


Fig. 12. Schermata del file di testo relativo alle concordanze della parola proverbio. Il cerchio mostra come viene evidenziato il nodo nel file .txt.

Come si può notare, l'allineamento del nodo viene ottenuto esclusivamente tramite l'utilizzo dei segni di maggiore e minore prima e dopo il nodo, il che facilita certamente l'individuazione della parola *proverbio*, anche se, non trattandosi di un vero allineamento della parola al centro della schermata, evidenziata per di più con un colore diverso rispetto al contesto, come succede lavorando direttamente con la schermata online (cfr. Fig. 2), l'occhio impiega un tempo maggiore per individuare la parola *proverbio*.

Vediamo in Fig. 13 come abbiamo proceduto alla sistematizzazione dei risultati e delle relative informazioni. Abbiamo deciso di disporre ogni informazione su una riga diversa. Abbiamo lasciato l'URL così come compariva nel file e poi abbiamo aggiunto l'etichetta di identificazione della tipologia testuale, seguita, nella riga successiva, dall'anno di pubblicazione del testo. A queste tre righe di informazioni abbiamo fatto seguire, sempre andando a capo, il contesto d'uso della parola *proverbio*. Abbiamo letto ogni contesto, non solo per individuare informazioni come la tipologia testuale e la datazione del testo, ma anche per capire se quell'occorrenza della parola *proverbio* doveva essere conservata oppure se ricadeva in una delle altre categorie di scarto. In seguito, per favorire il riconoscimento manuale dei diversi esempi ed evitare di avere tutte le righe attaccate, abbiamo lasciato una riga bianca tra un esempio e l'altro, riga che, se necessario, poteva comunque essere riempita con osservazioni che avrebbero composto una sorta di sezione dedicata alle note.

```

# Corpus: itWaC (Italian Web)
# Hits: 3706
# Query word,"proverbio" 3706

http://www.algherocronache.it/RisorseHTML/Archivio/collegamenti/2003/Calcio2003/Torres/TOR_2_03.htm
Articolo
2003
Bernardo Mereu pu√≤ tirare , dunque , un sospiro di sollievo , sebbene , come dice il < proverbio > fonte di saggezza , " una rondine non fa primavera " . La strada della salvezza √® ancora ardua : la penultima posizione in classifica in compropriet√† con il Sora , √® I√† ad attestare che non ci si pu√≤ cullare sugli allori di una pur altisonante vittoria (...).

http://www.amicigiornaleopinione.191.it/opinione/archivio/212/palazio2.html
Articolo
?
Mi ha guardato ridacchiando ... poi ha concluso : " S√† , ma c' √® anche un vecchio < proverbio > che dice : Non tutto il male vien per nuocere ! " . Ricordati di me Pi√π forte e con maggior convinzione di prima , il giovane regista romano racconta le aspirazioni soffocate e la vita complicata e deludente che appartiene a molti .

http://www.ioma.it/modules.php?name=News&file=article&sid=49&mode=thread&order=0&thold=0
Articolo
?
Lo hanno visto come un film concettuale , lontano dalla realt√† . In realt√† Sin city √® un film noir . " Benicio Del Toro a proposito della scena girata da Tarantino : " Io ero gi√† confuso avendo due registi . Ma esiste un < proverbio > che dice : due cuori sono meglio di uno , qualche volta tre sono pure meglio . E' stato divertente " . Clive Owen sulla collaborazione tra i due registi : " Frank ha realizzato un ambiente cos√† immaginifico che averlo sul set √® stato fantastico .

```

Fig. 13. Esempio di una schermata del file di testo in cui abbiamo sistematizzato le informazioni necessarie per ogni risultato ottenuto.

Come si osserva nell'esempio riportato qui sopra, alcuni caratteri speciali non venivano decodificati correttamente nel file di testo. Per questa ragione, tramite la funzione "trova e sostituisci", abbiamo proceduto alla correzione dei caratteri errati. Data la maggiore presenza di vocali accentate e il più ampio numero di accenti grafici utilizzati in francese, questi problemi sono stati riscontrati soprattutto nel corpus frWaC

Una volta ottenuto il file di testo con i soli esempi passati al vaglio, con le diverse informazioni necessarie su righe diverse e dopo aver sostituito i caratteri errati con quelli giusti, abbiamo importato tutti i dati in una pagina di Microsoft Word. Questo perché, per creare una tabella a partire da un file di solo testo, avremmo potuto sfruttare la possibilità che offre Word di convertire il testo in tabella. Abbiamo così impostato, come criterio di suddivisione delle diverse colonne il cosiddetto "segno di paragrafo": banalmente, il comando invio utilizzato per andare a capo. In questa maniera, siamo riusciti a creare una tabella che rispettava l'ordine delle informazioni da noi evidenziate, disponendole in colonne diverse. Una volta corretti alcuni errori (trattandosi di un lavoro manuale, poteva capitare che nella fase di aggiustamento del file di testo saltassimo alcune righe e così i contenuti non corrispondevano più alle colonne giuste), abbiamo esportato la tabella in Excel⁶⁶ per poter finalmente procedere all'analisi dei dati ottenuti. In Fig. 14 è possibile osservare la tabella in Excel composta sulla base della scrematura delle concordanze della parola *proverbio*, mentre in Fig. 15 è possibile vedere il file Excel relativo alle concordanze ottenute grazie al corpus frWaC.

⁶⁶ Abbiamo utilizzato la seguente versione: Microsoft® Excel® 2001 per Mac. Versione 14.6.4 (160422).

	A	B	C	D	E	F	G	H	I	J	K	L
	URL	TIPOLOGIA	DATA	CONTESTO								
1	http://www.cjargne.it/erwinmaier.htm	Articolo	?	La stima per l'immensa generosità, di cui era capace; la gratitudine per i gesti di amicizia, in particolare una visita all'ospedale, senza che ne avesse alcun obbligo, solo per amicizia. Un proverbio dice che i migliori ci lasciano presto; non è sempre vero, per fortuna, ma Erwin era certamente "uno dei migliori", una di quelle persone di cui abbiamo bisogno. E sono insostituibili.								
2	http://www.gazzettino.it/VisualizzaArticolo.php?Codice=2525105&Luogo=Vicenza&Data=2005-7-30&Pagina=BASSANO	Articolo	?	Tutto si è concluso con una risata collettiva, rimane il fatto che prima o poi si conierà un proverbio "dimmi che suoneria hai e ti dirò chi sei". Per il momento - soprattutto d'estate quando si sta più all'aperto - la varietà non è eccessiva, al pari della creatività.								
3	http://www.satyriconweb.it/articolo.asp?tipo=monselice&id_articolo=87	Articolo	?	In questa situazione non poca colpa ha l'ex pluriassessore Ghidotti. Chi troppo vuole nulla stringe, recita un vecchio proverbio. Come si sa gli antichi non sbagliano, e il timore di Ghidotti che un allargamento della maggioranza avrebbe portato ad un suo ridimensionamento (tra l'altro auspicabile, non								

Fig. 14. Pagina Excel relativa ai risultati convalidati dopo la precedente fase di pulitura del corpus itWaC.

	A	B	C	D	E	F	G	H	I	J	K
	URL	TIPOLOGIA	DATA	CONTESTO							
1	http://heaven-rpg.forumactif.fr/hopital-f38/une-chambre-isolee-t30230.htm	Littérature	?	(...) finalement, elle s'exclama un peu gênée : - Désolé pour la crise de nerfs mais, je n'aime pas quand on me plante sur place ^^ " Elle vit Galcyan éteindre une flamme et lui dit d'un ton moqueur. - On ne joue pas avec le feu ! Tu ne connais pas le proverbe - ^^ -							
2	http://www.animaux.fr/incident.php	Forum	?	je ne laisse jamais ma chienne seule avec des enfants ou errant sans surveillance, d'ailleurs, elle a pas le droit de s'éloigner de moi de plus de 5 mètres à la campagne !! J'ai opté pour le proverbe - " il vaut mieux prévenir que guérir " Si plus de personnes faisaient comme moi, il y aurait moins d'accidents stupides qui gachent la vie à pleins de gens !!							
3	http://paroisseschaville.free.fr/article.php?id_article=139	Article	?	Un proverbe - plutôt vrai ! - dit qu'on ne prête qu'aux riches. Toi, tu nous dis que le maître fait donner au plus riche le talent du 3 e serviteur, et que celui qui a, il lui sera donné plus encore.							
4	http://www.lire.fr/imprimer.asp/idC=52048	Article	?	(...) art du récit et d'une science des couleurs de bon aloi, mis au service d'une revisitation, pélemêle, de Gog et Magog, Sodome et Gomorrhe, sans oublier Adam et Eve. On attend avec curiosité la suite pour savoir si, en effet, comme le veut le proverbe - qui tombe pile -, « nul n'est							

Fig. 15. Pagina Excel relativa ai risultati convalidati dopo la precedente fase di pulitura del corpus frWaC.

3.7. Occorrenze dei proverbi introdotti dalla parola *proverbio*

Il primissimo dato che ci siamo apprestati a verificare, utile per capire se quanto avessimo fatto potesse essere valido dal punto di vista quantitativo e in un'ottica contrastiva tra i due corpora, era il numero di occorrenze della parola *proverbio* in italiano e della parola *proverbe* in francese.

Come si è già visto (cfr. § 3.3), il corpus itWaC, dopo opportuna scrematura dei risultati, restituisce 1.110 proverbi mentre il corpus frWaC, sempre dopo aver eliminato i risultati inutili o controproducenti ai fini della nostra ricerca, restituisce 974 proverbi. Per vedere se le grandezze dei proverbi ottenuti sono comparabili, facciamo una semplice proporzione su base 100 con il numero totale di occorrenze della parola *proverbio* in italiano e *proverbe* in francese. In italiano, la percentuale di proverbi ottenuti è del 30% circa, e in francese del 28% circa. Dato che la percentuale francese è molto simile a quella italiana, possiamo asserire con una ragionevole certezza che le due ricerche sono state condotte secondo criteri davvero comparabili. Le piccole differenze riscontrabili sono spiegabili con la diversa composizione dei due corpora e con differenze legate alle due diverse culture analizzate.

Appurato ciò, la prima osservazione operata consisteva nel capire se e quanti proverbi si ripetevano nelle diverse concordanze della parola *proverbio* e *proverbe*. Dopo aver esportato i nostri dati in Excel, abbiamo copiato, in una casella a parte, i proverbi di ogni concordanza, il che avrebbe evidenziato, appunto, eventuali ripetizioni di proverbi. Mentre eseguivamo questa fase, ci siamo resi conto dell'importanza anche di un altro aspetto legato all'enunciazione dei nostri proverbi, ovvero dei diversi *marcatori di proverbi*, seguendo la terminologia coniata nel capitolo precedente. In effetti, procedendo alla ricerca delle parola *proverbio* in contesto, siamo arrivati a determinare il nostro campione di proverbi da analizzare. Tuttavia, questo studio potrebbe servire anche a futuri progetti sviluppabili nello stesso ambito al confine tra paremiologia e *corpus linguistics*. Sarebbe interessante capire se è possibile trovare una gamma di proverbi utilizzando il metodo *corpus-driven* con marcatori più specifici rispetto alla sola parola *proverbio*, al fine di compiere ricerche più mirate che evitino di produrre così tanti risultati di scarto.

Pertanto, abbiamo deciso di evidenziare in grassetto non solo ogni proverbio ma anche tutti i marcatori di proverbi nella loro forma più estesa, per poi copiare separatamente i

marcatori e i proverbi in due colonne diverse. Mostriamo ora, in Fig. 16, questa fase di elaborazione dati relativa al corpus itWaC. Ricordiamo che lo stesso procedimento è stato applicato al corpus frWaC.

Per velocizzare questa fase, con un solo comando abbiamo contemporaneamente selezionato la formula che introduceva il proverbio e il proverbio stesso, mettendo le parole selezionate in grassetto e copiando tutta la stringa selezionata nelle due colonne subito a destra (nel nostro caso, si tratta delle colonne “E” ed “F”). Questo era il metodo più veloce per compiere tale procedimento. Ciononostante, come si può vedere in Fig. 16, i segni di punteggiatura rimanevano anche nelle due colonne aggiunte.

	A	B	C	D	E	F	G	H	I	J
	URL	TIPOLOGIA	DATA	CONTESTO	MARCATORE DI PROVERBIO	PROVERBIO				
1	http://www.cjargne.it/erwinmaier.htm	Articolo	?	La stima per l'immensa generosità, di cui era capace; la gratitudine per i gesti di amicizia, in particolare una visita all'ospedale, senza che ne avesse alcun obbligo, solo per amicizia. Un proverbio dice che i migliori ci lasciano presto ; non è sempre vero, per fortuna, ma Erwin era certamente "uno dei migliori", una di quelle persone di cui abbiamo bisogno. E sono insostituibili.	Un proverbio dice che	i migliori ci lasciano presto				
2	http://www.gazzettino.it/VisualizzaArticolo.php?Codice=2525105&Luogo=Vicenza&Data=2005-7-30&Pagina=BASSANO	Articolo	?	Tutto si è concluso con una risata collettiva, rimane il fatto che prima o poi si conierà un proverbio "dimmi che suoneria hai e ti dirò chi sei" . Per il momento - soprattutto d'estate quando si sta più all'aperto - la varietà non è eccessiva, al pari della creatività.	si conierà un proverbio	"dimmi che suoneria hai e ti dirò chi sei".				
3	http://www.satyriconweb.it/articolo.asp?tipo=monselice&id_articolo=87	Articolo	?	In questa situazione non poca colpa ha l'ex pluriassessore Ghidotti. Chi troppo vuole nulla stringe, recita un vecchio proverbio . Come si sa gli antichi non sbagliano, e il timore di Ghidotti che un allargamento della maggioranza avrebbe portato ad un suo ridimensionamento (tra l'altro auspicabile, non	recita un vecchio proverbio.	Chi troppo vuole nulla stringe,				

Fig. 16. Colonne (E ed F) introdotte per evidenziare la formula che introduce il proverbio e il proverbio stesso.

Si è scelto di utilizzare Excel come programma di gestione e sistematizzazione dei proverbi ottenuti, in quanto offre la possibilità di poter gestire parole o testi, a condizione che essi si trovino all'interno di una cosiddetta cella. Tuttavia, se si vuole procedere a studi statistici come calcolare il numero di occorrenze per ogni proverbio oppure stilare una lista di frequenza degli stessi proverbi, è necessario che il contenuto di due celle sia perfettamente identico perché il programma riconosca che si tratti davvero della stessa cosa. Ciò significa che, ad esempio, anche la sola presenza di una virgola in una cella contenente un proverbio esattamente uguale a quello di un'altra cella, ma che non contempla quella virgola, impedirà

al programma di capire che si tratta dello stesso proverbio. Non solo: anche la presenza di uno spazio in più oppure di maiuscole determinerà il non riconoscimento dell'identità delle celle da parte del programma.

In effetti, bisogna sottolineare che stiamo lavorando con materiale presente in Rete e prodotto dagli stessi utenti del Web: si tratta, pertanto, di materiale che può presentare errori di battitura, un'ortografia scorretta oppure diverse formattazioni a livello di caratteri maiuscoli e minuscoli. Il ricercatore deve, dunque, considerare questi elementi e affrontarli senza eliminarli (in quanto anche tali caratteristiche possono rivestire un'importanza notevole e fornire numerose informazioni, se non altro sul registro linguistico utilizzato e/o sull'estrazione sociale degli autori). È, quindi, importante mantenere tali elementi ma è altrettanto necessario trovare metodi per consentire al computer di evidenziare uguaglianze laddove ve ne siano.

Per prima cosa, abbiamo deciso di eliminare qualunque lettera maiuscola dai nostri proverbi. Contemporaneamente, sono stati tolti tutti i segni di punteggiatura, comprese le parentesi e le virgolette, nonché gli spazi doppi (ovvero due battiti consecutivi della barra spaziatrice, errore molto frequente che non permette il corretto riconoscimento di uguaglianze da parte del programma). Per eliminare le maiuscole, abbiamo utilizzato la funzione “=MINUSC” di Excel che permette di eliminare completamente tutte le maiuscole. In seguito, al fine di eliminare i segni di punteggiatura, le parentesi, le virgolette e i doppi spazi, ci siamo serviti della funzione “Trova e sostituisci”: in questi casi, nella casella “trova” abbiamo inserito i caratteri che dovevamo sostituire, oppure i doppi spazi, e nella casella “sostituisci” non abbiamo inserito alcunché oppure solo uno spazio a seconda dei casi e delle nostre esigenze. In questa maniera siamo riusciti a eliminare automaticamente i caratteri indesiderati.

Per non perdere informazioni preziose e faticosamente ottenute, abbiamo aggiunto un foglio alla pagina di Excel in cui abbiamo copiato le colonne che avevamo fino allora compilato (cfr. Fig. 16). Nel nuovo foglio di Excel, chiamato “Proverbi senza punteggiatura”, abbiamo poi proceduto alla rimozione tutti i segni di “disturbo” di cui abbiamo parlato poc'anzi. Ecco, dunque, in Fig. 17, il nuovo file Excel comprendente il primo foglio di proverbi con punteggiatura e il secondo foglio di proverbi senza punteggiatura (la “punteggiatura” è un'etichetta convenzionale da intendersi in senso lato).

	A	B	C	D	E	F	G	H	I	J
	URL	TIPOLOGIA	DATA	CONTESTO	MARCATORE DI PROVERBIO	PROVERBI senza punteggiatura solo minuscole senza parentesi				
1	http://www.cjargne.it/erwinmaier.htm	Articolo	?	La stima per l'immensa generosità, di cui era capace; la gratitudine per i gesti di amicizia, in particolare una visita all'ospedale, solo per amicizia. Un proverbio dice che i migliori ci lasciano presto; non è sempre vero, per fortuna, ma Erwin era certamente "uno dei migliori", una di quelle persone di cui abbiamo bisogno. E sono insostituibili.	Un proverbio dice che	i migliori ci lasciano presto				
2	http://www.gazzettino.it/VisualizzaArticolo.php?Codice=2525105&Luogo=Vicenza&Data=2005-7-30&Pagina=BASSANO	Articolo	?	Tutto si è concluso con una risata collettiva, rimane il fatto che prima o poi si conierà un proverbio "dimmi che suoneria hai e ti dirò chi sei". Per il momento - soprattutto d'estate quando si sta più all'aperto - la varietà non è eccessiva, al pari della creatività.	si conierà un proverbio	dimmi che suoneria hai e ti dirò chi sei				
3	http://www.satyronweb.it/articolo.asp?tipo=monsellece&id=1444	Articolo	?	In questa situazione non poca colpa ha l'ex pluriassessore Ghidotti. Chi troppo vuole nulla stringe, recita un vecchio proverbio. Come si sa gli antichi non sbagliano, e il timore di Ghidotti che un allargamento della maggioranza avrebbe portato ad un suo	recita un vecchio proverbio.	chi troppo vuole nulla stringe				

Fig. 17. Nuovo foglio chiamato "Proverbi senza punteggiatura".

Abbiamo cominciato a notare che c'erano vari proverbi che si ripetevano, anche se ve n'erano alcuni diversi ma che potevano essere considerati varianti di altri proverbi. Per una maggiore chiarezza, abbiamo stilato una prima lista di proverbi non ripetuti, il che significa che abbiamo chiesto a Excel di elencare i valori univoci, ossia solamente le celle diverse. In altre parole, anche se avessimo avuto quattro celle identiche, tre di queste sarebbero state nascoste e solo una visualizzata. Questa operazione è stata possibile tramite il seguente procedimento: una volta selezionata la colonna per la quale si volevano visualizzare solo i valori univoci, abbiamo cliccato su "Dati", "Filtro avanzato" e selezionato "Copia in un'altra posizione". Abbiamo così inserito, nel campo "Copia in:" il riferimento della prima cella in alto di una colonna vuota. Infine, abbiamo spuntato la casella "Copia univoca dei record" e abbiamo cliccato su "OK". In questa maniera, nella colonna vuota selezionata, abbiamo ricavato una lista di valori univoci dei nostri proverbi. Abbiamo così esaminato manualmente tutti i proverbi ottenuti alla ricerca di varianti, in maniera da uniformare il più possibile i nostri risultati e stilare, così, una lista di frequenza dei nostri proverbi.

Tuttavia, ci siamo ben presto trovati di fronte a un problema considerevole: una volta trovate alcune varianti di altri proverbi, avremmo comunque dovuto scegliere un'unica variante. Quale metodologia adottare per operare tale scelta?

Dato che l'etichetta che avremmo dovuto assegnare a questi proverbi sarebbe stata solamente un'etichetta formale utile al programma per calcolare il numero di volte in cui si ripetevano proverbi dello stesso tipo, abbiamo deciso di servirci di dizionari di proverbi.

Abbiamo già parlato, in § 2.9.1, dei dizionari di proverbi italiani utilizzati per il primo tentativo, per altro fallito, di costituire il campione di proverbi da cercare in itWaC. Si esporranno, ora, le regole stabilite per il reperimento di varianti di proverbi nei dizionari, necessarie per avere la certezza di trattare nella stessa maniera il materiale tratto da itWaC e quello tratto da frWaC.

Per prima cosa, abbiamo scelto un ordine di consultazione delle opere, in base alla praticità di consultazione delle stesse, intesa come la capacità dei dizionari di offrire diverse metodologie di ricerca dei proverbi stessi, oppure in base alla più alta probabilità di reperire i proverbi in nostro possesso. Sono stati, così, prediletti quei dizionari che prevedevano indici secondari contenenti, ad esempio, tutte le parole presenti nei proverbi e non solamente le parole chiave. Sappiamo bene, infatti, che i dizionari di proverbi scelgono spesso una parola chiave per ogni proverbio, solitamente quella più rilevante, in base alla quale categorizzare i proverbi o disporli in ordine alfabetico. Tuttavia, bisogna tenere in considerazione che molto spesso è proprio quella parola chiave che viene modificata al momento dell'enunciazione dei proverbi, il che riduce le possibilità di ritrovare quel proverbio. Ecco perché è di fondamentale importanza avere anche altri indici che contemplino tutte le parole dei proverbi (o almeno quelle più semanticamente rilevanti, come sostantivi, verbi, aggettivi e a volte perfino avverbi⁶⁷).

In base a queste considerazioni, abbiamo deciso di scegliere come primo dizionario *Il grande dizionario dei proverbi italiani* di Guazzotti e Oddera (2006). La ragione di tale scelta è da attribuire alla praticità di utilizzo dello stesso dizionario: esso, infatti, ordina dapprima i proverbi in ordine alfabetico escludendo i soli articoli determinativi e indeterminativi iniziali (pertanto, se un proverbio comincia con un articolo determinativo dovrà essere cercato a partire dalla prima parola utile: ad esempio, il proverbio *La gatta frettolosa fece i gattini ciechi*, andrà cercato in base alla parola *gatta*; qualora *un, un', una, uno* siano numerali e non

⁶⁷ Ad esempio, *Il grande dizionario dei proverbi italiani* di Guazzotti e Oddera offre in appendice un nutrito "Indice delle parole contenute nei proverbi" (2006: 641-991) con cui si possono reperire proverbi a seconda delle principali parole che in essi appaiono. Prendiamo ad esempio il proverbio *Chi va piano va sano e va lontano*. Questo si può reperire grazie a tutte le seguenti parole: *piano* (utilizzato qui in funzione di avverbio); *andare* (verbo); *sano* (utilizzato qui in funzione di avverbio); *lontano* (utilizzato qui in funzione di avverbio). Il pronome *chi* non è presente in questo elenco.

più articoli, se si trovano all'inizio del proverbio possono essere utilizzati per cercare il proverbio stesso⁶⁸); successivamente, offre un indice tematico in cui dispone i proverbi a seconda della categoria tematica a cui sono stati assegnati (sebbene sia forse l'indice meno fruibile); infine, propone un indice delle parole presenti nei proverbi (cfr. nota 67 per osservare il funzionamento di tale indice) in cui sono elencati, sotto ogni parola, i proverbi in cui essa è contenuta. Diversamente da altri dizionari che offrono lo stesso tipo di indice, il Guazzotti e Oddera (2006) non fornisce semplicemente il riferimento numerico del relativo proverbio, ma indica già nell'indice il proverbio in forma estesa: in questa maniera, per reperire il proverbio nel dizionario, è sufficiente cercare il proverbio riportato nell'indice.

Nel caso in cui un proverbio non fosse presente in questo dizionario, abbiamo deciso di consultare, come secondo dizionario, il *Dizionario dei proverbi italiani* di Lapucci (2006). Questo dizionario differisce dal Guazzotti e Oddera (2006) per il metodo scelto nell'ordinamento dei proverbi: essi sono certamente disposti in ordine alfabetico, ma non a partire dalla prima parola utile del proverbio stesso, bensì sulla base della parola chiave di ogni proverbio, che solitamente è la parola più emblematica del proverbio stesso. Ciononostante, come si diceva in precedenza, non sempre è possibile reperire un proverbio sulla base di una sola parola. Questa circostanza spiega forse la presenza di un altro indice in cui sono elencate tutte le parole principali dei proverbi presenti nel dizionario. Rispetto al dizionario di Guazzotti e Oddera (2006), in questo indice non sono riportati i proverbi nella forma in cui li troviamo all'interno del dizionario, ma vengono semplicemente indicati, per ogni parola, i numeri identificativi dei proverbi che a loro volta contengono quella parola. Dato che il reperimento di proverbi con questo dizionario potrebbe richiedere un tempo maggiore rispetto a quello consentito dal dizionario Guazzotti e Oddera (2006), abbiamo preferito utilizzare il Lapucci (2006) nel caso in cui i proverbi non fossero presenti nel primo dizionario.

Abbiamo già parlato del dizionario di Boggione e Massobrio (2007) in § 2.9.1. Oltre ad avere la particolarità di inserire anche proverbi dialettali o locali che, invece, rivestono uno scarsissimo interesse nei confronti del nostro studio, questo dizionario ordina i propri proverbi secondo grandi categorie tematiche. Ciò significa che il tempo necessario per reperire i proverbi desiderati si allunga di molto, in quanto si deve procedere per tentativi, non sapendo a priori in quale categoria siano disposti i proverbi. Inoltre, in caso di insuccesso, non si ha la

⁶⁸ Per tutte le regole di consultazione del dizionario, si veda Guazzotti e Oddera (2006: 7).

certezza, come per gli altri due dizionari, di non trovare il proverbio desiderato perché effettivamente esso non è attestato nel dizionario: si potrebbe presentare il caso in cui sia impossibile trovare un proverbio semplicemente perché è stato categorizzato sotto un'area tematica a cui l'utente potrebbe non aver pensato. Per tutte queste ragioni, abbiamo deciso di non consultare affatto, in questa fase, il dizionario di Boggione e Massobrio (2007).

Per quanto riguarda i dizionari di proverbi francesi⁶⁹, abbiamo riscontrato una minore adeguatezza ai fini del nostro lavoro rispetto ai dizionari italiani. Anzitutto, sorprendono i numeri: i dizionari italiani più comuni e facilmente reperibili non riportano meno di 11.000 proverbi (Guazzotti e Oddera, 2006), ma possono anche più che raddoppiare tale cifra e toccare i 25.000 proverbi (Lapucci, 2006); per i dizionari francesi i numeri sono ben diversi. Il dizionario di Dournon (1986), che pure sembra abbastanza completo, raccoglie 3.500 proverbi di Francia, mentre il dizionario di Maloux (2009) raccoglie 10.000 tra proverbi, sentenze e massime, ma non solo di Francia, bensì di tutto il mondo, comprendendo anche proverbi, sentenze e massime del mondo classico. Infine, il dizionario di Montreynaud et al. (1989) presenta un totale di 5.736 proverbi francesi e di Francia suddivisi in 2.136 *proverbes français*, 863 *dictons météorologiques* (francesi), 1.277 *dictons de la croyance* (francesi) e 1460 *proverbes locaux et historiques* (sempre francesi). In questo elenco non includiamo quelli che vengono definiti *proverbes du monde*, i quali costituiscono un'ampia sezione del dizionario: dato che ci interessiamo solo ai proverbi francesi e di Francia, e posto che in questo dizionario i proverbi del resto del mondo sono separati dai primi, li abbiamo esclusi in partenza da questo calcolo.

Dopo aver illustrato i tre dizionari di cui ci saremmo serviti, abbiamo deciso l'ordine di consultazione: dapprima, ci saremmo concentrati sul dizionario di Dournon (1986), poi su quello di Maloux (2009) e, infine, su quello di Montreynaud et al. (1989).

Rispetto ai tre dizionari, il Dournon (1986) era quello di più pratico utilizzo: anzitutto, perché esso considera esclusivamente i proverbi francesi e di Francia e, in secondo luogo, poiché dispone i proverbi in ordine alfabetico a seconda della parola chiave di ogni proverbio, solitamente quella più importante del proverbio. Questo metodo, però, non è sempre efficace,

⁶⁹ I dizionari di proverbi francesi sono stati cercati in base a quanto farebbe un ipotetico utente italiano che si trovi in questa situazione: pertanto, abbiamo cercato nelle principali biblioteche pubbliche dell'Emilia-Romagna (in quanto per consultare un dizionario è necessario recarsi nella biblioteca che lo conserva, essendo quasi sempre vietato il prestito, compreso il cosiddetto prestito interbibliotecario) ma anche in Internet tra i principali siti di vendita online di libri, come Amazon (<<https://www.amazon.it/>>) e Abebooks (<<http://www.abebooks.it/>>).

in quanto non di rado può accadere che l'utente consideri fondamentale una parola che non viene ritenuta tale dall'autore del dizionario. Per ovviare a questo problema, prima di considerare un proverbio non attestato nel dizionario è necessario controllare se le altre parole del proverbio siano o meno presenti nel dizionario. Ricordiamo, infine, che in Dournon (1986) non è presente alcun indice parallelo che permetta altre piste di ricerca.

Nel caso in cui in Dournon (1986) non fossero presenti i proverbi cercati, abbiamo deciso di consultare come seconda opera il Maloux (2009). Questo dizionario propone un indice finale in cui sono riportati i proverbi attestati nel dizionario a seconda delle diverse parole che li compongono, il che aiuta notevolmente il ricercatore. Ciononostante, questo dizionario è stato utilizzato come seconda scelta, in quanto accanto ai proverbi francesi di Francia considera proverbi, sentenze e massime di tutto il mondo.

È capitato, nel corso di questa fase di omologazione delle varianti dei proverbi, che i primi due dizionari non bastassero. Ecco perché abbiamo consultato anche il Montreynaud et al. (1989), lasciandolo tuttavia come ultima scelta in quanto suddivide da subito i proverbi in diverse categorie (ovvero: *proverbes*, *dictons météorologiques*, *dictons de la croyance* e *proverbes locaux et historiques*), le quali sono a loro volta suddivise in numerose categorie e sottocategorie tematiche. Abbiamo notato che, anche in questo caso, la ricerca di proverbi sulla base di categorie tematiche è poco affidabile e richiede un tempo maggiore. Questo dizionario propone altresì un indice delle parole contenute nei proverbi, nel quale, tuttavia, sono presenti le parole abbinata ai codici dei relativi proverbi: ciò significa che per reperire il proverbio desiderato può essere necessario un lasso di tempo alquanto dilatato.

Dopo aver stabilito l'ordine di utilizzo dei dizionari, abbiamo formulato l'ennesima regola: siccome anche all'interno dei dizionari, i proverbi vengono presentati assieme ad altre varianti, abbiamo pensato di scegliere quello che definiamo *proverbio standard*, ovvero la forma principale del proverbio che solitamente si trova al primo posto nel dizionario oppure la forma sotto la quale sono disposte le altre varianti o, ancora, la prima in ordine alfabetico.

Cosa fare, invece, nel caso di un'eventuale impossibilità nel reperire i nostri proverbi all'interno dei dizionari? Ebbene, in questo caso abbiamo utilizzato il corpus stesso per reperire la variante più frequente del proverbio, scegliendo proprio quella. In altri termini, abbiamo cercato il proverbio stesso nel corpus, omettendo la parola *proverbio* e ottenendone,

così, le varie occorrenze. In alternativa, abbiamo scelto di cercare questi proverbi in Internet, tramite il diffuso motore di ricerca Google⁷⁰.

Dopo aver stilato le principali norme da seguire, abbiamo effettuato il riconoscimento manuale delle varianti dei nostri proverbi, nonché il reperimento della forma standard per uniformare i nostri risultati. Puntualizziamo, per sgombrare il campo da eventuali fraintendimenti, che questo lavoro è stato effettuato solamente nel caso in cui fossero attestate, tra i nostri risultati, due o più varianti diverse dello stesso proverbio. Per tutti i casi in cui i proverbi non venivano ripetuti oppure in cui le varianti erano le medesime (ovvero compariva un solo proverbio di quel tipo nella nostra lista di valori univoci), non abbiamo proceduto alla ricerca degli stessi nei dizionari. La ragione di questo sta in un'ottica di risparmio di tempo e risorse: dato che la ricerca della forma standard ha senso solo nei casi in cui sono presenti più varianti, per mere ragioni legate ai limiti della macchina, è inutile estendere la ricerca a risultati che non presentano varianti dissimili.

Già in questa fase, è emersa una caratteristica di notevole importanza per tutto il nostro lavoro: la presenza di proverbi modificati. Dato che non era ancora giunto il momento di analizzare il fenomeno delle modifiche dei proverbi, abbiamo deciso di trattarlo alla stregua delle varianti dei proverbi.

In seguito a queste riflessioni, abbiamo proceduto alla fase di ricerca di varianti. In Fig. 18 è visibile la nostra schermata di Excel aggiornata con le successive colonne relative alla ricerca di varianti di proverbi (mostriamo solamente il file italiano, ricordando che quello francese è identico). Spieghiamo brevemente come abbiamo costruito le colonne aggiuntive. Anzitutto, abbiamo creato *ex novo* una colonna (nella nostra immagine è la colonna G) in cui abbiamo copiato tutti i proverbi che avevamo già individuato. In questo modo, abbiamo tenuto traccia dei proverbi originali, sebbene avessero già attraversato la fase di eliminazione di punteggiatura e lettere maiuscole. Questa nuova colonna è stata intitolata “Proverbi di colonna a sx (copiati): con più varianti si sceglie quella attestata nel G.D.P.I.”, dove l'acronimo “G.D.P.I.” sta per *il Grande Dizionario dei Proverbi Italiani* di Guazzotti e Oddera (2006). Nel caso in cui un proverbio avesse avuto una occorrenza oppure una sola variante, lo stesso proverbio sarebbe stato copiato nella colonna F. Nel caso in cui, invece, fosse stato necessario reperire la forma standard per un proverbio presente con diverse

⁷⁰ URL: <<https://www.google.it/>>. © 2015 Google Inc. Tutti i diritti riservati. Google e il logo Google sono marchi registrati di Google Inc.

varianti nel corpus, avremmo cercato la forma standard nel dizionario di Guazzotti e Oddera (2006). Una volta trovata, l'avremmo inserita nella cella della stessa riga nella colonna H, intitolata "Come si trova in G.D.P.I.", e nella colonna G, dal titolo "Proverbi di colonna a sx (...)". Questo passaggio serviva per tenere traccia delle forme standard cercate sui dizionari e di quei proverbi che, invece, non necessitavano di questo tipo di trattamento. Infatti, se una cella della colonna "Come si trova in G.D.P.I." è vuota, ciò significa che il proverbio non è stato cercato in alcun dizionario per le ragioni appena menzionate. Per segnalare, invece, una fonte diversa rispetto al primo dizionario usato, avremmo utilizzato la colonna H, ma specificando tra parentesi la fonte stessa.

	A	B	C	D	E	F	G	H	I	J	K	L
	URL	TIPOLOGIA	DATA	CONTESTO	MARCATORE DI PROVERBIO	PROVERBI senza punteggiatura solo minuscole senza parentesi	PROVERBI DI COLONNA A SX (copiati) con più varianti si sceglie quella attestata nel G.D.P.I.	COME SI TROVA IN G.D.P.I.	PROVERBI SIMILI IN G.D.P.I.	ATTESTAZIONI NEL WEB		
1	http://www.ejargne.it/erwinmaier.htm	Articolo	?	La stima per l'immensa generosità, di cui era capace; la gratitudine per i gesti di amicizia, in particolare una visita all'ospedale, senza che ne avesse alcun obbligo, solo per amicizia. Un proverbio dice che i migliori ci lasciano presto; non è sempre vero, per fortuna, ma Erwin era certamente "uno dei migliori", una di quelle persone di cui abbiamo bisogno. E sono insostituibili.	Un proverbio dice che	i migliori ci lasciano presto	sono sempre i migliori quelli che se ne vanno	non attestato				
2	http://www.gazzettino.it/VisualizzaArticolo.php?Codice=2525105&Luogo=Vicenza&Data=2005-7-30&Pagina=BAS SANO	Articolo	?	Tutto si è concluso con una risata collettiva, rimane il fatto che prima o poi si conierà un proverbio "dimmi che suoneria hai e ti dirò chi sei". Per il momento - soprattutto d'estate quando si sta più all'aperto - la varietà non è eccessiva, al pari della creatività.	si conierà un proverbio	dimmi che suoneria hai e ti dirò chi sei	dimmi con chi vai e ti dirò chi sei	dimmi con chi vai e ti dirò chi sei				
3	http://www.satyriconweb.it/articolo.asp?tipo=monselice&id_articolo=87	Articolo	?	In questa situazione non poca colpa ha l'ex plurassessore Ghidotti. Chi troppo vuole nulla stringe, recita un vecchio proverbio. Come si sa gli antichi non sbagliano, e il timore di Ghidotti che un allargamento della maggioranza avrebbe portato ad un suo ridimensionamento (tra l'altro auspicabile, non certo però in favore di un leghista) ha finito per essere la condanna di (...).	recita un vecchio proverbio.	chi troppo vuole nulla stringe	chi troppo vuole nulla stringe	chi troppo vuole nulla stringe				

Fig. 18. Tabella Excel dopo la fase di ricerca della forma standard attestata nei dizionari.

Successivamente, abbiamo ritenuto indispensabile creare la colonna I, "Proverbi simili in G.D.P.I.", utilizzata per quei casi in cui avessimo trovato proverbi affini a quelli cercati, benché abbastanza diversi dal punto di vista lessicale e sintattico. Anche in questo caso, se avessimo reperito un proverbio simile in un'altra fonte, lo avremmo indicato tra parentesi subito dopo il proverbio stesso. Trattandosi di proverbi abbastanza diversi, avremmo comunque optato per la variante più comune all'interno del corpus.

Per contro, se avessimo dovuto reperire un proverbio in Rete in quanto non attestato nei dizionari e poco frequente nel nostro corpus, avremmo utilizzato la colonna J e riportato quella variante sempre nella colonna G.

Grazie a questa tappa, abbiamo potuto trovare confronti di proverbi e ridurre al massimo il numero di quelli che abbiamo definito “valori univoci”, ovvero di forme standard di proverbi senza considerarne le varianti. Per terminare questa fase, ci mancava ancora un passaggio: stilare una lista di frequenza dei nostri proverbi, ovvero quantificare la frequenza di ognuno dei nostri valori univoci. Per fare ciò, abbiamo utilizzato la funzione di Excel “=CONTA.SE”, procedendo come illustrato qui di seguito.

Per prima cosa, abbiamo creato un nuovo file Excel in cui abbiamo importato i dati della colonna G del file iniziale che comprendeva le varianti di proverbi raggruppate secondo la forma standard reperita nei dizionari. A questo punto, abbiamo copiato i valori univoci di questa colonna in un'altra colonna (seguendo il procedimento spiegato poco sopra). A questo punto abbiamo utilizzato la funzione “=CONTA.SE”: basta indicare l'intervallo di dati della colonna contenente tutti i proverbi, compresi quelli ripetuti, facendo attenzione che prima della lettera della colonna e del numero di riga ci sia il simbolo \$, ed esplicitare il riferimento di una cella appartenente alla colonna dei valori univoci, perché il software conti le volte in cui quella cella è presente nell'intervallo di dati selezionato. Facciamo un esempio. Nel nuovo file, i nostri proverbi vanno dalla cella “A2” alla cella “A1111”. La colonna dei valori univoci è la colonna C, mentre vogliamo che nella colonna D compaia il numero di occorrenze dei proverbi nella colonna C. Per fare ciò, digiteremo la seguente formula nella colonna D: “=CONTA.SE(\$A\$2:\$A\$1111;D2)”. Così facendo il sistema andrà a cercare quante volte si ripete il proverbio presente in D2 nell'intervallo da A2 ad A1111. Trascinando questa formula fino alla cella della colonna D, che corrispondeva all'ultimo proverbio della colonna C, abbiamo ottenuto automaticamente la lista di frequenza dei nostri proverbi. Infine, abbiamo ordinato i proverbi a seconda del numero delle loro occorrenze, dal più frequente al meno frequente. In caso di parità di frequenza, abbiamo deciso di ordinare i proverbi alfabeticamente.

	A	B	C	D	E	F	G
	PROVERBI DI COLONNA A SX (copiati): con più varianti si sceglie quella attestata nel G.D.P.I.		ELENCO VALORI UNIVOCI	OCCORRENZE			
1	[i fatti che] sono argomenti dalla testa dura		non c'è due senza tre	16			
2	30 giorni ha novembre con aprile * giugno * e settembre		aiutati che Dio t'aiuta	13			
3	a buon intenditor poche parole		l'unione fa la forza	13			
4	a buon intenditor poche parole		chi ben comincia è a metà dell'opera	11			
5	a carnevale ogni scherzo vale		tra il dire e il fare c'è di mezzo il mare	11			
6	a caval donato non si guarda in bocca		una mela al giorno leva il medico di torno	11			
7	a caval donato non si guarda in bocca		chi va piano va sano e va lontano	10			
8	a caval donato non si guarda in bocca		il lupo perde il pelo ma non il vizio	9			
9	a caval donato non si guarda in bocca		non tutto il male vien per nuocere	9			
10	a caval donato non si guarda in bocca		chi lascia la via vecchia per la nuova, sa quel che lascia e non sa quel che trova	8			
11	a chi nasce sfortunato piove sul didietro anche se sta seduto		chi trova un amico, trova un tesoro	8			
12	a chi tutto a chi niente		l'abito non fa il monaco	8			
13	a forza di tenere aperta la bocca si sono chiuse le orecchie		la gatta frettolosa fece i gattini ciechi	8			
14	a giovane eremita vecchio diavolo		la mamma dei cretini è sempre incinta	8			
15	a mali estremi estremi rimedi		sbagliando s'impara	8			
16	a nemico che fugge ponti d'oro		chi rompe paga e i cocci sono suoi	7			
17	a nemico che fugge ponti d'oro		chi si contenta gode	7			
18	a pensar male si fa peccato ma si indovina		errare è umano, perseverare è diabolico	7			
19	a pensar male si fa peccato ma si indovina		il buon giorno si vede dal mattino	7			
20	a pensar male si fa peccato ma si indovina		l'occasione fa l'uomo ladro	7			
21							

Fig. 19. Lista di valori univoci dei proverbi e relative occorrenze.

Vediamo dunque, in Fig. 20 e in Fig. 21, la distribuzione delle occorrenze dei proverbi reperiti tramite i marcatori *proverbio* e *proverbe*. Sottolineiamo che il numero totale di valori univoci ottenuti è di 617 per itWaC e di 630 per frWaC. Si nota da subito una schiacciante maggioranza relativa ai proverbi aventi una sola occorrenza. Per il corpus itWaC, abbiamo 449 proverbi con frequenza 1, che totalizzano il 73% del totale dei proverbi reperiti mediante la modalità descritta, contro 518 proverbi aventi frequenza 1 in frWaC, che rappresentano l'82% dei risultati ottenuti. Seguono le occorrenze 2, che rappresentano rispettivamente il 12% e il 7% dei risultati ottenuti in itWaC e in frWaC. Dall'occorrenza 3, che si attesta al 4% sia per itWaC sia per frWaC, si va sempre scendendo fino ad arrivare allo 0,2% delle prime occorrenze per quantità di proverbi, ossia l'occorrenza 16 per itWaC e l'occorrenza 14 per frWac.

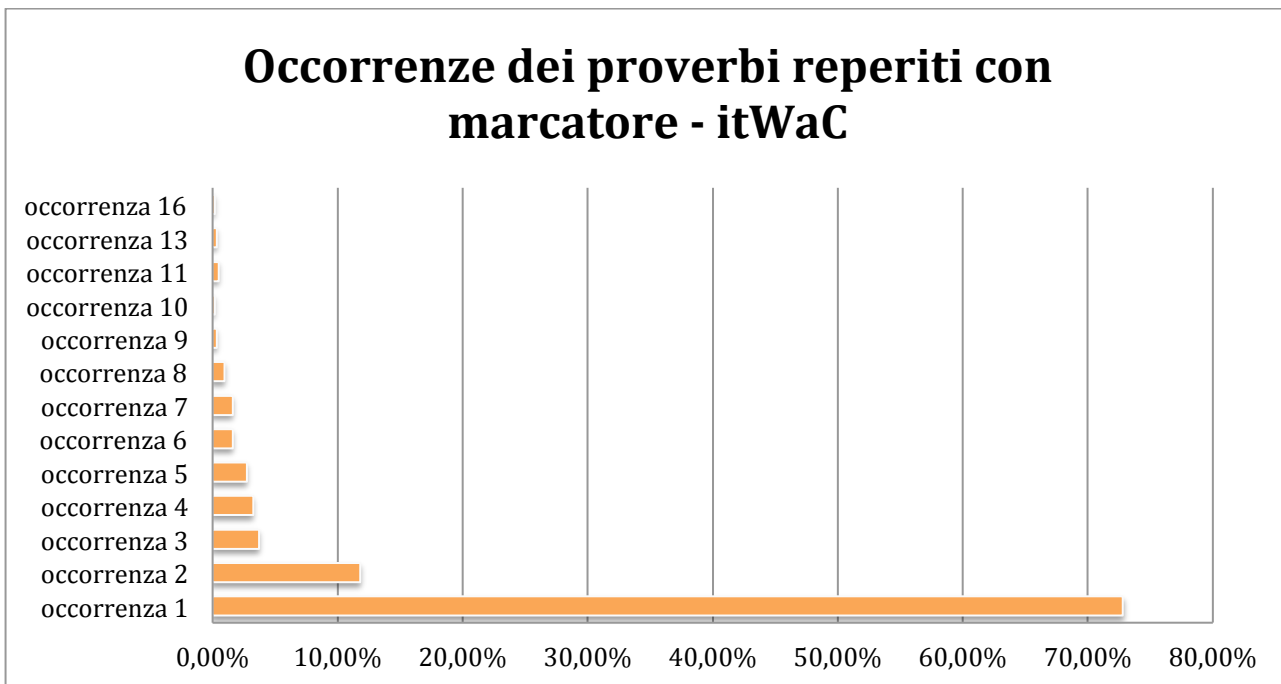


Fig. 20. Occorrenze dei proverbi individuati grazie alla parola *proverbio* in itWaC.

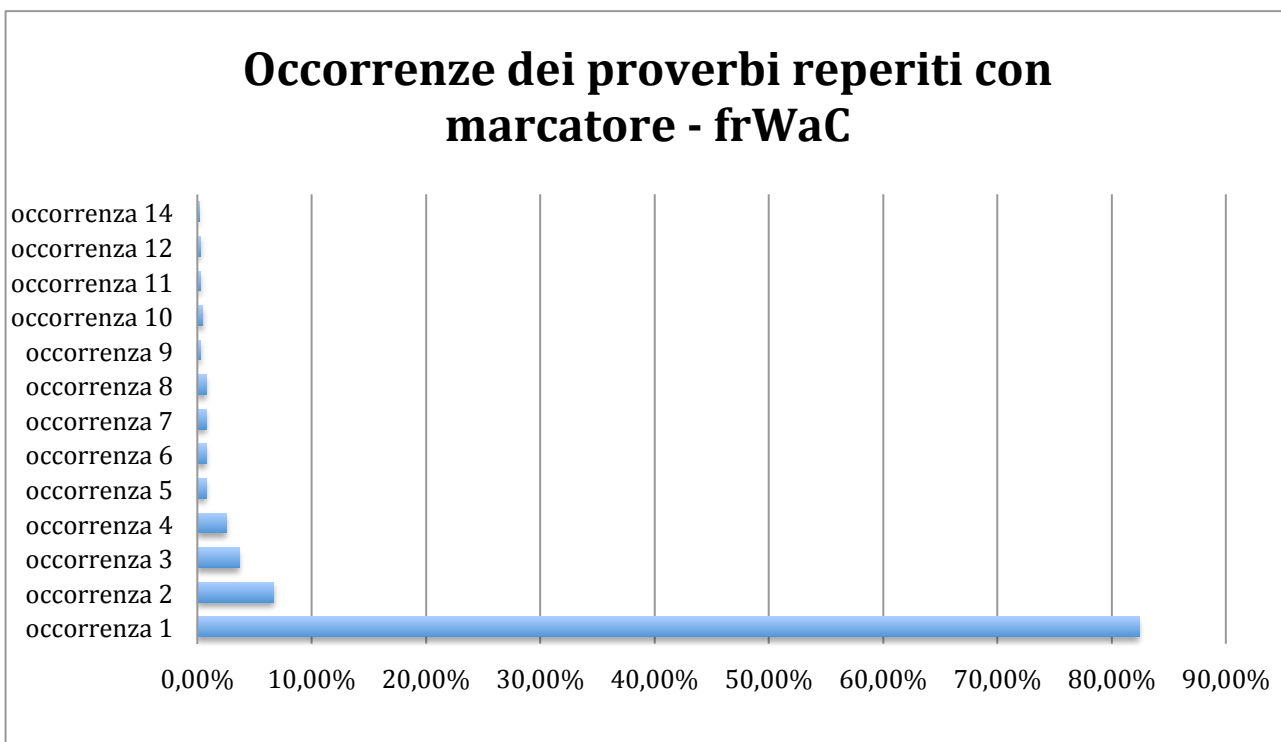


Fig. 21. Occorrenze dei proverbi individuati grazie alla parola *proverbe* in frWaC.

3.8. Liste di frequenza di proverbi: i frutti del nostro lavoro

I frutti principali della ricerca esposta nel presente capitolo sono probabilmente le liste di frequenza dei proverbi ottenute tramite l'approccio *corpus-driven* a partire dai corpora itWaC e frWaC e seguendo le tappe delineate in questo capitolo.

Dapprima, abbiamo spiegato come discernere tra proverbi utili ai fini della presente ricerca e proverbi meno interessanti, in quanto stranieri o locali, non più attuali oppure legati a intenti metaparemiologici piuttosto che a finalità comunicative.

Abbiamo così analizzato i dati derivanti da questa scrematura e, in particolare, le informazioni estrapolate dai proverbi non italiani e non francesi: questa fase è stata compiuta sulla base esclusiva dei *marcatori di proverbio* da noi utilizzati, vale a dire la parola *proverbio* e *proverbe*. Il fatto che i marcatori di proverbio possano permettere al ricercatore non solo di reperire proverbi ma anche di ottenere grandi quantità di informazioni sui proverbi enunciati, ribadisce la rilevanza e la validità della metodologia individuata.

Successivamente alla fase di scrematura dei dati, siamo arrivati a costruire la nostra banca dati contenente i proverbi da noi ottenuti tanto in italiano quanto in francese, corredati di informazioni molto utili, come l'URL degli esempi, la tipologia testuale in cui essi ricorrevano, la datazione degli stessi e il loro contesto d'uso. Con questa banca dati le strade di ricerca percorribili sono davvero molte. Alcune di queste verranno già esposte nel prossimo capitolo, altre invece verranno solo indicate in attesa di ulteriori studi in questo ambito.

Come si diceva, il frutto della ricerca esposta nel presente capitolo risiede probabilmente nelle liste di frequenza dei proverbi individuati in itWaC e in frWaC, tramite la ricerca dei marcatori di proverbi, e consultabili nell'Allegato 1 e nell'Allegato 2.

Desideriamo, tuttavia, terminare questo capitolo con alcuni dati. Grazie alla metodologia di reperimento semiautomatico di proverbi, tramite i principi dell'approccio *corpus-driven*, siamo riusciti a ottenere ben 617 proverbi non ripetuti (i cosiddetti valori univoci) da itWaC e 630 proverbi da frWaC. Molti di questi proverbi ricorrono il numero di volte indicato a fianco di ciascuno nelle liste dei due allegati menzionati. Si tratta di dati di grandissimo interesse, in quanto si riferiscono ai proverbi italiani e francesi introdotti dalle parole *proverbio* e *proverbe*, il che ci deve fare pensare che i proverbi utilizzati nella comunicazione odierna siano molti di più.

Questo ha un grandissimo valore non solo per la presente ricerca, ma anche per la paremiologia nel suo insieme: i dati reali dimostrano che l'opinione diffusa, persino in ambito accademico, secondo la quale il proverbio non rivesta alcun interesse ai fini linguistici è errata.

Nel prossimo capitolo, si proseguirà nell'analisi dei dati emersi, fornendo ulteriori prove a supporto dell'importanza del proverbio.

4. Analisi di proverbi italiani e francesi

Il lavoro esposto nel terzo capitolo, relativo all'estrazione semiautomatica dei proverbi da itWaC e da frWaC, è di estrema importanza per approfondire l'analisi dei proverbi, che viene ora proposta.

In questo capitolo, pertanto, cercheremo di capire il grado di affidabilità del marcatore di proverbio da noi scelto e i tipi di risultati a cui permette di giungere. Vedremo, così, quali caratteristiche hanno i proverbi introdotti dalla parola *proverbio*, in italiano, e *proverbe*, in francese, nonché se la prima fase di ricerca effettuata tramite la concordanza di questi due marcatori sia in linea con lo studio dei proverbi reperiti senza alcun marcatore.

Vedremo, successivamente, che le modifiche dei proverbi rivestono un grande interesse, in quanto rappresentano un fenomeno ampiamente attestato nei corpora utilizzati. Questo punto ci darà la possibilità di esaminare un tema molto dibattuto in ambito paremiologico: quello sulla fissità dei proverbi. Ci chiederemo se il tema della fissità dei proverbi sia davvero infondato oppure se, al contrario, essi possano essere considerati fissi.

Anzitutto, vediamo come sono stati cercati, nei due corpora, i proverbi precedentemente ottenuti senza utilizzare alcun marcatore di proverbio.

4.1. Scelta dei proverbi da cercare e analizzare in itWaC e frWaC

Quanto sono indicativi i marcatori di proverbi utilizzati per cercare in maniera semiautomatica i proverbi nei nostri due corpora? Non è certo facile rispondere a questa domanda, non solo perché non abbiamo molti termini di paragone con cui confrontare la nostra analisi, ma anche perché le variabili da prendere in considerazione sono davvero molte.

Per questa ragione, ci concentreremo, *in primis*, sulla seconda fase della nostra ricerca, vale a dire sul reperimento dei proverbi già individuati grazie alle concordanze dei marcatori di proverbio. In altre parole, una volta ottenuto il campione di proverbi da analizzare (elencati negli allegati 1 e 2), questi sarebbero stati cercati nei due corpora senza bisogno di aggiungere la parola *proverbio* o *proverbe* ai criteri di ricerca.

Nel capitolo precedente, sono già stati presentati alcuni dati importanti sulle occorrenze dei proverbi individuati soltanto sulla base di un marcatore di proverbi comune, ovvero la parola *proverbio* e la parola *proverbe*. Si è notato che la stragrande maggioranza degli esempi reperiti rientra, tanto per il corpus italiano quanto per quello francese, in quella che abbiamo definito "occorrenza 1". Ciò significa che i tre quarti dei risultati ottenuti sono

stati enunciati, assieme alla parola proverbio, da un solo utente della Rete. Questo fatto può essere problematico perché ci restituisce esempi che condividono la caratteristica di contenere almeno una frase definita proverbio dal parlante, ma che non trovano riscontro nell'occorrenza dello stesso proverbio enunciato in altri contesti. La linguistica dei corpora permette la formulazione di ipotesi sulla base di fatti rilevanti, in quanto ricorrenti e ripetuti. È chiaro che, con fenomeni che ricorrono una sola volta, tale presupposto viene meno.

Per questa ragione, siccome nulla assicura che i proverbi aventi occorrenza 1 siano effettivamente riconosciuti come tali dall'intera comunità dei parlanti, abbiamo deciso di non prendere in considerazione l'intera categoria per la seconda fase della nostra ricerca. I proverbi da cercare nei corpora sarebbero dunque stati scelti tra quelli a partire dall'occorrenza 2.

Per determinare il campione di proverbi da cercare nei due corpora, ci siamo ispirati alla soglia minima indicata da Sinclair (2005) per formulare una descrizione lessicografica di una parola: tale limite minimo è di almeno 20 occorrenze. Ora, siamo consapevoli dell'incommensurabilità di due grandezze così diverse come le parole e le frasi, pertanto non stiamo affermando che questo valore debba essere valido anche per i proverbi. Tuttavia, abbiamo deciso di applicare tale criterio anche all'analisi dei proverbi per una serie di ragioni. In primo luogo, perché dovevamo banalmente trovare un criterio per restringere il nostro campo di analisi e focalizzarci sugli elementi più rappresentativi, ossia più frequenti. In secondo luogo, perché dovevamo individuare una soglia minima al di sotto della quale non poter studiare i proverbi sia per la parte relativa all'italiano sia per quella riguardante il francese. Infine, perché, sulla base di alcuni tentativi compiuti su alcuni proverbi scelti casualmente, la soglia di 20 occorrenze ci è sembrata tutto sommato ragionevole.

Siamo partiti da una considerazione: anche l'occorrenza 2 non ci sembrava sufficiente a produrre considerazioni attendibili sui proverbi. Un'occorrenza uguale a due significava che, nella migliore delle ipotesi, due utenti avevano definito una certa frase proverbio e, nella peggiore delle ipotesi, che lo stesso utente aveva utilizzato la stessa espressione in due testi diversi, oppure che si trattava di duplicati che non eravamo riusciti a eliminare nella fase di scrematura dei corpora. Inoltre, avremmo preferito considerare i proverbi che avessero almeno una frequenza d'uso pari a 3, in quanto l'idea che si potesse triangolare un fatto linguistico così complesso come il proverbio ci sembrava essenziale per garantire una certa affidabilità alle nostre osservazioni. Abbiamo così preso campioni casuali rappresentanti il

25% dei proverbi con frequenza 2 e 3 e siamo andati a vedere quante volte si ripetevano nei nostri due corpora senza che fossero accompagnati da alcun marcatore. Sapevamo bene che allargando le maglie della nostra ricerca, ovvero diminuendo le parole da cercare oppure introducendo alcune categorie morfosintattiche al posto di parole precise, avremmo incluso tra i risultati le versioni modificate dei nostri proverbi. Tuttavia, siccome non avevamo ancora preso in considerazione il tema delle variazioni e delle modifiche dei proverbi, abbiamo deciso, come regola, di cercare solamente le versioni attestate inizialmente in itWaC, senza voler andare a cercare varianti non attestate nel nostro primo campione di proverbi.

Abbiamo scelto, generando numeri casuali, i seguenti 18 proverbi italiani aventi 2 occorrenze.

Proverbi con occorrenza 2	Occorrenze reali in itWaC
a buon intenditor poche parole	47
agosto moglie mia non ti conosco	3
a nemico che fugge ponti d'oro	3
al cuor non si comanda	84
anche l'occhio vuol la sua parte	135
casa dolce casa	24
chi primo arriva meglio alloggia	9
chi si è scottato con l'acqua calda ha paura anche di quella fredda	1
chi va al mulino s'infarina	5
del senno di poi son piene le fosse	14
il gioco non vale la candela	131
il tempo è galantuomo	66
la notte porta consiglio	40
la vendetta è un piatto che va gustato freddo	13
mentre i medici discutono il malato muore	2
sacco vuoto non sta in piedi	4
tale il padre tale il figlio	9

una mano lava l'altra e tutt'e due lavano il viso	49
---	----

Fig. 22. Campione del 25% scelto casualmente tra i proverbi con frequenza iniziale pari a 2 e relative occorrenze reali in itWaC. Si noterà che nei proverbi sono assenti segni di punteggiatura e maiuscole. Oltre a queste modifiche non sono stati apportati altri cambiamenti ai proverbi stessi.

Sui 18 proverbi scelti casualmente, 8 hanno un'occorrenza reale inferiore a 20. Ciò significa che meno della metà dei nostri proverbi aventi due occorrenze, ossia il 44%, ha una frequenza reale superiore a 20. Vediamo i dati relativi ai proverbi di frequenza iniziale pari a 3 (Fig. 23).

Proverbi con occorrenza 3	Occorrenze reali in itWaC
chi è causa del suo mal pianga se stesso	30
il silenzio è d'oro	62
la curiosità uccise il gatto	2
lontano dagli occhi lontano dal cuore	13
non si getta via l'acqua sporca col bambino dentro	6
sono sempre i migliori quelli che se ne vanno	19

Fig. 23. Campione del 25% scelto casualmente tra i proverbi con frequenza iniziale pari a 3 e relative occorrenze reali in itWaC. Si noterà che nei proverbi sono assenti segni di punteggiatura e maiuscole. Oltre a queste modifiche non sono stati apportati altri cambiamenti ai proverbi stessi.

In questo caso, il 50% dei proverbi mostra un'occorrenza reale maggiore o uguale alle 20 unità. In realtà, è con i proverbi di occorrenza 4 che riusciamo a ottenere una percentuale superiore al 50% (cfr. Fig. 24): tale percentuale si attesta all'80%.

Proverbi con occorrenza 4	Occorrenze reali in itWaC
chi di spada ferisce di spada perisce	44
il diavolo fa le pentole ma non i coperchi	92
non si può avere la botte piena e la	243

moglie ubriaca	
non vendere la pelle dell'orso prima di averlo preso	15
patti chiari, amicizia lunga	52

Fig. 24. Campione del 25% scelto casualmente tra i proverbi con frequenza iniziale pari a 4 e relative occorrenze reali in itWaC.

È stata individuata, in questa maniera, la soglia di occorrenza iniziale (ossia relativa ai proverbi reperiti grazie al marcatore di proverbio) minima per avere una probabilità superiore al 50% di trovare gli stessi proverbi almeno 20 volte nell'intero corpus. Come riprova, abbiamo eseguito gli stessi calcoli sui proverbi aventi un'occorrenza iniziale maggiore o uguale a 5, non solo in italiano ma anche in francese. Presentiamo, in Fig. 25, un riepilogo delle probabilità percentuali di reperire più di 20 proverbi nell'intero corpus sulla base delle occorrenze degli stessi proverbi accompagnati dalla parola *proverbio* e dalla parola *proverbe*.

Occorrenze con marcatori	Probabilità in % di avere occorrenze in itWaC > 20	Probabilità in % di avere occorrenze in frWaC > 20
2	44%	17%
3	50%	67%
4	80%	50%
5	100%	100%
6	100%	60%
7	100%	100%
8	100%	100%
9	100%	100%
10	100%	100%
11	100%	100%
13 (itWaC) / 12 (frWaC)	100%	100%
16 (itWaC) / 14 (frWaC)	100%	100%

Fig. 25. Probabilità che i proverbi italiani e francesi delle diverse occorrenze iniziali indicate nella prima colonna abbiano nei due corpora una frequenza d'uso maggiore o uguale a 20.

Nel prossimo grafico (Fig. 26) abbiamo cercato di sintetizzare schematicamente questo andamento, mettendo a confronto nuovamente i dati italiani con quelli francesi, al fine

di individuare il punto di incremento della probabilità di trovare proverbi con occorrenza reale maggiore o uguale a 20 e le occorrenze in cui tale percentuale si stabilizza sul 100%.

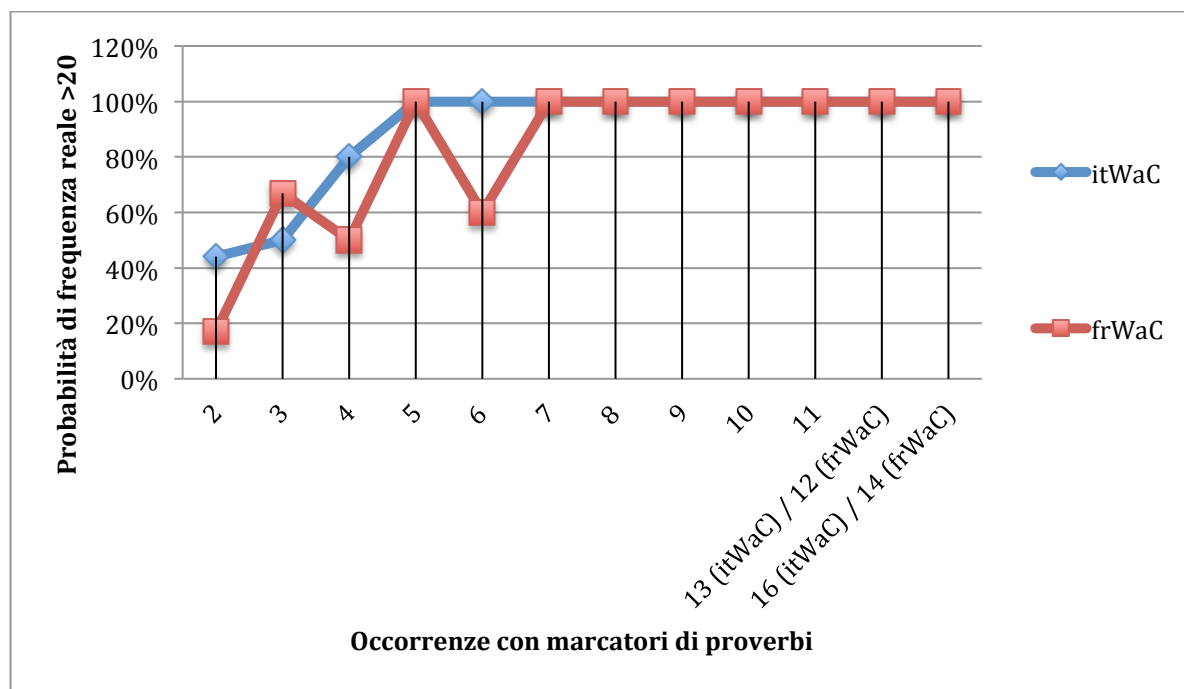


Fig. 26. Andamento della probabilità di trovare più di 20 proverbi nei corpora itWaC e frWaC in funzione delle occorrenze determinate in base ai relativi marcatori di proverbi.

Grazie al grafico appena riportato, notiamo un elemento molto importante, che è stato determinante per farci scegliere la frequenza minima iniziale da cui partire per cercare i proverbi nei due corpora di riferimento. Vediamo che se la curva dei proverbi di itWaC è più lineare, uniforme e presenta un'unica tendenza chiara all'aumento, la curva dei proverbi francesi è più ondulatoria e meno lineare. Vi è, tuttavia, un punto in comune tra i proverbi italiani e quelli francesi che può essere considerato la soglia minima al di sotto della quale non si può andare per cercare i proverbi: questa è rappresentata dall'occorrenza 4. Se in italiano vi era già una tendenza all'aumento, anche se questa si attesta al di sopra del 50% di probabilità solo a partire dalla frequenza 4, in francese tale tendenza è interrotta dalla flessione registrata in corrispondenza della frequenza 6. Tuttavia, se si escludesse l'occorrenza 5 di frWaC, che arriva probabilmente troppo presto a una probabilità del 100%, o se si facesse comunque una media tra i valori delle occorrenze 4, 5 e 6, si avrebbe una tendenza molto più lineare. Ad ogni modo, a partire dall'occorrenza 4, non si notano più, nei

due corpora, probabilità inferiori al 50% di trovare più di 20 occorrenze. Da questo ragionamento deriva la scelta di prendere in considerazione solamente i proverbi aventi un'occorrenza iniziale maggiore o uguale a 4.

4.2. Come cercare i proverbi scelti nei due corpora

Una volta risolto il problema legato al campione di proverbi da utilizzare per compiere ulteriori ricerche nei due corpora, ci siamo chiesti come effettuare tali ricerche nella pratica. Durante la fase di costruzione del nostro corpus di proverbi individuati tramite la parola *proverbio* e *proverbe*, si era già notata la presenza di proverbi modificati, non tanto sulla base di nostre conoscenze paremiologiche o di dizionari di proverbi, quanto in virtù di certe espressioni che introducevano alcuni proverbi modificati. Si considerino i seguenti esempi:

- (22) (...) per parafrasare un proverbio che non mi ricordo esattamente: chi di esercizio colpisce di esercizio perisce.⁷¹
- (23) Parafrasando il famoso proverbio, potremmo dire che l'abito fa il coniglietto.⁷²
- (24) Parafrasando un proverbio... non c'è 7 senza 8!!!!!!⁷³
- (25) Ho sempre cercato di fare da me quello che molti cercano di fare in tre (derivato dal proverbio) e ho avuto spessissimo dei vantaggi rispetto alle azioni di gruppo.⁷⁴
- (26) Chissà, forse fra poco, aggiornando un antico proverbio molto popolare, potremmo affermare festanti che un lupino al giorno toglie l'asma di torno!⁷⁵
- (27) Sono proprio questi i momenti in cui si pensa a chi non c'e', che a dispetto del proverbio, ha sempre ragione.⁷⁶
- (28) Ciao Andrea, premetto che tutti possono sbagliare, premetto che al contrario di ciò che dice il proverbio esistono anche coloro che predicano male e razzolano bene, (...).⁷⁷
- (29) Bella serata e belli i coccetti che, contrariamente a quanto dice il proverbio, erano tuoi anche se non li rompevi.⁷⁸

⁷¹ URL: <twiki.dsi.uniroma1.it>.

⁷² URL: <fumetto.leonardo.it>.

⁷³ URL: <http://www.cicloweb.it/forum/viewthread.php?fid=4&tid=2099&action=printable>.

⁷⁴ URL: <http://www.aduc.it/dyn/dilatua/dila_mostra.php?id=113141>.

⁷⁵ URL: <http://www.ticaebio.it/giornalino/notiziaarch.asp?id=921>.

⁷⁶ URL: <http://www2.raisport.rai.it/news/sport/nazionale/200005/27/392fd79e02066/>. L'apostrofo al posto del carattere è viene impiegato nel testo originale. Come sempre, non abbiamo modificato gli esempi riportati.

⁷⁷ URL: <http://www.cicloweb.it/forum/viewthread.php?tid=609>.

⁷⁸ URL: <http://www.bloggers.it/rainking/>.

- (30) Il proverbio dice più o meno così: «Dai nemici mi guardi Iddio che dagli amici mi guardo io».⁷⁹

Notiamo una costante valida per tutti questi proverbi: se consideriamo come marcatore di proverbio non solo la parola *proverbio* ma anche tutta l'espressione in cui tale parola si trova, notiamo che ci sono elementi impiegati dal parlante per sottolineare che non si tratta di proverbi comuni, bensì di proverbi modificati e non citati nella loro forma standard. Ci sono dei marcatori che indicano chiaramente questo status di proverbio modificato: tipicamente, il verbo *parafrasare* che troviamo negli esempi (22), (23) e (24) altro non fa se non segnalare che è stata prodotta una modifica nell'enunciazione del proverbio. Anche l'espressione *derivato dal proverbio*, che troviamo nell'esempio (25), e *aggiornando un antico proverbio*, presente in (26), hanno la stessa funzione. In (28) e in (29) si segnala chiaramente una posizione di disaccordo con quanto asserito dal proverbio nella sua forma standard, il che spiega le modifiche apportate al proverbio stesso, mentre in (30) l'autore ammette di non ricordarsi com'è esattamente il proverbio, cosa che lascia intendere che il proverbio sia effettivamente diverso rispetto alla forma standard. Con tutti i marcatori di proverbi come quelli appena esposti, è quasi scontato trovare proverbi modificati.

Vediamo, ora, qualche esempio reperito nel corpus frWaC.

- (31) Cette prévention-là consiste à informer par avance, pour sensibiliser aux actions préventives. Parodiant un autre proverbe, on pourrait dire que deux préventions valent mieux qu'une.⁸⁰
- (32) Mais, paraphrasant le proverbe, rappelons que si les intérêts mesquins et parfois inavouables aboient, la caravane de la recherche passe et suit son chemin en toute sérénité, car elle n'a de comptes à rendre qu'à l'histoire.⁸¹
- (33) Paraphrasant un proverbe bien connu, je dirais que la "liberté de la presse" s'arrête où commence "la liberté de tout citoyen à disposer de sa propre vie privée".⁸²

⁷⁹ URL: <http://www.dsmilano.it/html/Pressroom/2003/06/uni3_0608_tranfaglia-dove-sono-i-movimenti.htm>.

⁸⁰ URL: <<http://bulletin.conseil-national.medecin.fr/Archives/html/308/308BOMN308P13A1.htm>>.

⁸¹

URL: <http://ww2.editionsladecouverte.fr/webcc/sog_dec/notice_reference.html?F_ean13=9782707149619&X_param=EP>.

⁸² URL: <<http://lapresseaujourd-aujourd-hui.blog.20minutes.fr/archive/2007/10/12/divorce-ou-not-divorce.html>>.

- (34) L’habit ne fait pas le moine et je vais inventer un proverbe à moi la djellaba ne fait pas le musulman.⁸³
- (35) (...) mais ça fait : "..... à qui sait attendre". Bon je sais pas trop ce qu’obtient celui qui sait attendre vu que je me souviens plus du début du proverbe !!⁸⁴
- (36) Certains fumeurs sont ainsi passés de deux paquets à cinq "vrais" cigarettes par jour. Une aide au sevrage qui fait très "tendance" et qui prouve que le proverbe est faux : il peut y avoir de la fumée sans feu ! bonne année à tous !⁸⁵
- (37) Tahar, consterné s’écrit : Mi, kis ti fou ? Djamel lui répond : Toi, ti prends les roues. Moi, j’i prends la radio. !. !. !. !... De là, le proverbe : “Mieux vaut Tahar que Djamel”.⁸⁶

Negli esempi (31), (32), (33) e (34), i verbi *parodier*, *paraphraser* e *inventer*, riferiti ai proverbi che li seguono, indicano in partenza che essi verranno citati in forma modificata. In (35) il parlante ammette di non ricordare bene l’inizio del proverbio, il che giustifica il fatto che effettivamente esso venga citato solo per metà (anziché *Tout vient à point à qui sait attendre* si dice solamente *à qui sait attendre*). In (36) l’espressione che introduce il proverbio indica già la probabilità di incontrare il proverbio stesso modificato, in quanto si sostiene che il proverbio standard “*est faux*”: tale disaccordo è, appunto, presente anche a livello di enunciazione del proverbio stesso che, anziché essere *Il n’y a pas de fumée sans feu*, diventa *Il peut y avoir de la fumée sans feu*.

In (37), invece, non ci sono etichette formali che indichino che il proverbio sia stato modificato. Tuttavia, la modifica era così evidente e così geniale che abbiamo deciso di inserire anche questo esempio. Si tratta di un caso di omofonia, in cui il proverbio, se pronunciato ad alta voce, sembra quasi identico al proverbio nella forma standard ma, effettivamente, ha una forma così irriconoscibile da renderne impossibile il ritrovamento senza conoscerne a priori la nuova forma. Si è scelto questo esempio anche per sottolineare che molto spesso, tanto in francese quanto in italiano, gli utenti della Rete si sbagliano a scrivere i proverbi in maniera più o meno volontaria: si può trattare, infatti, di errori

⁸³ URL: <http://carpediem.typepad.fr/carpe_diem/2006/11/a_labri.html>.

⁸⁴ URL: <<http://memecamouille.cowblog.fr/commentaires-2084646.html>>.

⁸⁵ URL: <<http://blog-notre-epoque-tf1.lci.fr/15-categorie-53293.html>>.

⁸⁶ URL: <<http://les-zhoms.easyforum.fr/humour-insolite-expression-hot-f9/proverbe-arabe-t1305.htm>>.

ortografici oppure di abbreviazioni tipiche del linguaggio delle *chat*. Si osservino gli esempi seguenti.

- (38) oué c vré on né sorti ensemble je c po combien de foi mé il i a un proverbe ki di il fo se kité souven pour aimer tjs !!!⁸⁷ [trascrizione secondo l'ortografia francese corretta: *Ouais, c'est vrai, on est sorti ensemble je ne sais pas combien de fois mais il y a un proverbe qui dit: il faut se quitter souvent pour (s') aimer toujours !*]
- (39) j ai 1 proverbe pour ton blog si sa tinteraisse "ne quitte jamais celui ki t aime pour celui ke tu aime. car celui ke tu aime te quittera pour celui kil aime"⁸⁸ [trascrizione secondo l'ortografia francese corretta: *J'ai un proverbe pour ton blog si ça t'intéresse : « Ne quitte jamais celui qui t'aime pour celui que tu aimes car celui que tu aimes te quittera pour celui qu'il aime »*]
- (40) Il est évident que chaque cas est unique et plus ou moins grave mais comme dit le proverbe : "vaut mieux prévenir que guérir".⁸⁹

I tre esempi illustrati indicano che, pur ammettendo che un proverbio non sia intenzionalmente modificato e sia, quindi, citato nella forma standard (almeno quella conosciuta dal parlante), non è detto che possa essere reperito se cercato secondo quella variante, in quanto scritto violando le più comuni norme ortografiche (in (38) e (39) viene utilizzato il linguaggio degli sms e delle *chat*) o grammaticali (notiamo in (40) la mancanza del soggetto obbligatorio per i verbi impersonali, tipica dell'oralità). Tutte queste considerazioni ci hanno posto di fronte all'ennesima scelta da compiere per trovare un *modus operandi* facilmente individuabile e determinabile, nonché valido e applicabile sia per il corpus italiano sia per quello francese.

Dato che i nostri proverbi avevano una frequenza iniziale di almeno 4 occorrenze, e si trovavano quindi in almeno quattro contesti diversi, abbiamo deciso di considerare fin da subito le modifiche presenti nei proverbi in nostro possesso. Per ognuno dei nostri proverbi, avremmo tentato di trovare una struttura che potesse essere, da un lato, abbastanza generale da reperire il maggior numero di occorrenze per ogni proverbio senza distinzione tra forme

⁸⁷ URL: <<http://love7.myblog.fr/alex-p28320.html>>.

⁸⁸ URL: <http://poemes_dtt.myblog.fr/la-semaine-p81587.html>.

⁸⁹ URL: <<http://oceane3664.skyrock.com/>>.

standard, varianti o proverbi modificati, ma, dall'altro lato, abbastanza selettiva da prendere in considerazione solamente i proverbi e non qualsiasi frase libera.

Per chiarire questo punto, facciamo alcuni esempi prendendo alcuni proverbi di diverse occorrenze iniziali. Precisiamo che d'ora in poi, quando parleremo di varianti, intenderemo proverbi aventi qualsiasi differenza a livello lessicale e morfosintattico, escludendo, però, eventuali disuguaglianze ortografiche e di punteggiatura.

Iniziamo col proverbio *Patti chiari e amicizia lunga*. Notiamo che, nella fase iniziale dello studio, questo proverbio ha occorrenza pari a 4 e presenta tre varianti principali:

- (41) a. Patti chiari, amicizia lunga / Patti chiari ed amicizia lunga
- b. L'amicizia è lunga quando i patti sono chiari
- c. Patti oscuri, amicizia corta

La domanda che ci dobbiamo porre è la seguente: qual è il denominatore comune di queste tre varianti? Come possiamo, con un'unica ricerca, reperire il maggior numero di varianti evitando di prendere in considerazione anche frasi libere che nulla hanno a che vedere con il proverbio stesso? Per rispondere a questi interrogativi, dobbiamo concentrarci sulle strutture che rimangono costanti. Riprendiamo le tre varianti, evidenziando in grassetto ciò che rimane invariato:

- (42) a. **Patti chiari, amicizia lunga / Patti chiari ed amicizia lunga**
- b. **L'amicizia è lunga quando i **patti** sono chiari**
- c. **Patti oscuri, amicizia corta**

Dal punto di vista lessicale, gli elementi costanti sono i sostantivi *patti* e *amicizia*. Dal punto di vista morfosintattico, osserviamo in due casi su tre la presenza dell'aggettivo *chiari* dopo il sostantivo *patti* e l'aggettivo *lunga*, posto dopo il sostantivo *amicizia*. Nel terzo caso, vi sono due aggettivi che mantengono la stessa posizione (seguono entrambi i sostantivi *patti* e *amicizia*), ma sono diversi dal punto di vista semantico (sono, rispettivamente, gli antonimi di *chiari* e *lunga*, vale a dire *oscuri* e *corta*). Facendo astrazione di tali regolarità lessicali e morfosintattiche, abbiamo pensato che fossero i due sostantivi presenti nella totalità degli esempi a essere necessari per ritrovare il proverbio a seconda delle sue diverse varianti e delle

versioni modificate, in quanto, in tutti gli esempi a nostra disposizione, erano i due sostantivi a rimanere invariati. Per questa ragione, abbiamo immesso nella stringa di ricerca i sostantivi *patti* e *amicizia*, dando la possibilità al programma di ricerca di considerare uno spazio da 1 a 5 parole di qualsiasi tipo tra i due sostantivi, al fine di comprendere anche quei risultati, simili alla frase (42) b, in cui le due parole sono intervallate da altri vocaboli, ivi compresi i segni di punteggiatura. Ecco in (43) la nostra ricerca: i lemmi virgolettati devono essere cercati tali e quali, mentre il codice $\{x,y\}$, dove x e y sono due valori numerici, indica che tra due parole ci devono essere da un minimo di “x” parole a un massimo di “y” parole:

(43) "patti" $\{1,5\}$ "amicizia"

Si ottengono in questa maniera 105 risultati (che diventano 93 escludendo 12 risultati ripetuti). Tuttavia, tra questi risultati figurano numerosi esempi (ben 44) in cui non è presente il proverbio cercato, come quelli riportati in (44):

- (44) a. (...) una volta essere adolescente significava non scendere mai a **patti**, significava credere nell'**amicizia** (...).
- b. (...) ora con uno, ora con l'altro paese, e molto spesso formando **patti** d'alleanza e di **amicizia** con uno mentre in gran segreto ne firmava un altro con il nemico (...).
- c. Regolamento per la stipula e gestione **patti** di gemellaggio, di **amicizia** e di fratellanza.
- d. Ad oggi sono oltre 300 i **patti** di gemellaggio e di **amicizia** stipulati nella nostra regione con città di tutti i paesi del mondo.

Ci si rende conto, così, che una ricerca di questo tipo non può essere accettabile per un lavoro come il nostro, in quanto invaliderebbe i risultati ottenuti, non concentrandosi solo su proverbi ma comprendendo anche frasi libere. Una probabile ragione di questo fenomeno potrebbe risiedere nel fatto che, contrariamente a quanto succede negli esempi (42), i due sostantivi cercati non sono seguiti da alcun aggettivo. Effettivamente, benché in uno dei nostri esempi gli aggettivi fossero diversi dal punto di vista semantico e lessicale, essi rimanevano invariati dal punto di vista morfosintattico. Questa potrebbe essere l'altra regolarità utile alla

ricerca del proverbio stesso. Per questa ragione, abbiamo introdotto l'etichetta "aggettivo" nella nostra ricerca, dove il codice [tag="A"] significa, appunto, qualsiasi parola, in quella precisa posizione, che sia stata etichettata morfosintatticamente come aggettivo. Abbiamo così modificato la precedente ricerca come segue:

(45) "patti" [tag="A"] []{1,5} "amicizia" [tag="A"]

In questa maniera otteniamo un numero inferiore di risultati (solamente 48 escluse le ripetizioni) rispetto a quelli ottenuti con la precedente ricerca (ben 93 che diventano, però, 49 escludendo i 44 risultati non pertinenti), ma abbiamo la certezza di ricavare esclusivamente esempi in cui è presente il proverbio desiderato e le sue varianti. L'unica variante che non riusciamo a trovare e che avevamo trovato con la ricerca precedente mostrata in (43) è la seguente: *Se i patti saranno chiari l'amicizia sarà lunga*⁹⁰. Questo è dovuto alla presenza del verbo *essere* tra il sostantivo *amicizia* e l'aggettivo *lunga*. Tuttavia, si tratta di un sacrificio necessario, in quanto, per trovare una sola ulteriore occorrenza del proverbio, sarebbe controproducente trovare 44 esempi in più composti da frasi libere, che nulla hanno in comune con il proverbio stesso.

Applichiamo, ora, lo stesso procedimento a un altro proverbio, questa volta di frequenza iniziale pari a 11 ed estremamente noto in italiano come *Tra il dire e il fare c'è di mezzo il mare*. Osserviamo, in (46), le diverse varianti del proverbio ottenute in seguito alla prima ricerca effettuata con il marcatore *proverbio*:

- (46) a. Tra / fra il dire e / ed il fare c'è di mezzo il mare
b. Tra il dire e il fare c'è di mezzo il classico mare
c. Tra il dire e il fare
d. C'è di mezzo il mare

Il primo aspetto da osservare è che, anche quando la variante è la stessa, ci possono essere delle differenze che riguardano, ad esempio, la scelta di *tra* al posto di *fra* e viceversa, oppure di *ed* al posto della più corretta congiunzione *e* senza la cosiddetta *d* eufonica. Vediamo, inoltre, che il 50% delle varianti è composto da versioni ridotte del proverbio, in cui

⁹⁰ URL: <<http://www.uil.it/uilscuola/uilscuolaonline/stampa/2000/091000A.htm>>.

si cita il proverbio parzialmente. Inoltre, si osserva, nella variante del punto *b*, l'introduzione di un aggettivo, *classico*, prima del sostantivo *mare*.

Abbiamo, così, cercato negli esempi gli elementi costanti e invariabili, sebbene gli esempi al punto *c* e *d* siano inconciliabili: in *c* si cita solamente la prima parte del proverbio e in *d* soltanto la seconda. Per redimere questi casi, si è scelto di non prendere in considerazione un *pattern* poco rappresentativo. Al contrario, nel caso in cui i vari *pattern* evidenziati nella prima fase di ricerca di proverbi fossero ugualmente presenti e rilevanti, si è deciso di prenderli tutti in considerazione, effettuando ricerche complementari ma distinte.

Tornando al nostro esempio (46), il *pattern* probabilmente meno rappresentativo è quello al punto *c*, ossia *Tra il dire e il fare*: in effetti, l'aggettivo *classico* segnala che la tipicità del proverbio sta proprio nella seconda parte. Inoltre, notiamo che l'aggiunta di questo aggettivo indica che è la seconda parte a essere più spesso modificata, il che suggerisce che è probabilmente la colonna portante dell'intero proverbio: infatti, benché venga modificata, assicura sempre un corretto riconoscimento del proverbio. Per queste ragioni, abbiamo cercato solo la seconda parte del proverbio, quella a nostro avviso più importante, seguendo le chiavi di ricerca di (47):

(47) "di" "mezzo" "il" []{0,3} "mare"

Tale ricerca fa emergere 213 risultati, escluse le ripetizioni. A riprova di questa scelta, vediamo che cosa sarebbe emerso se avessimo scelto di cercare solamente la prima parte del proverbio:

(48) ".*ra" "il" "dire" "e.*" "il" "fare"

Con questo tipo di ricerca, otteniamo quasi il doppio dei risultati rispetto a quella precedente, ma lo svantaggio è che molti esempi ricavati si distaccano dal proverbio e sconfinano nella frase libera, come quelli riportati in (49).

- (49) a. (...) ad evidenziare altre omissioni dell'esecutivo che rivelano un'incoerenza **tra il dire e il fare**, un comportamento verso le Forze di polizia e le Forze armate che sta provocando (...).⁹¹
- b. La differenza **tra il dire ed il fare**, tra i principi e la realtà.⁹²
- c. Il progetto **Tra il dire e il fare**: le attività di tirocinio come opportunità per la scuola.⁹³
- d. (...) il modello costruito risulta essere di aiuto per avvicinare i "luoghi" della riflessione formativa ed organizzativa ai "luoghi" di lavoro, accorciando lo spazio **fra il dire ed il fare**.⁹⁴

Tutto questo non accade con la ricerca evidenziata in (47), con la quale si cerca il proverbio sulla base della sua seconda parte, ammettendo delle modifiche anche in essa (come l'aggiunta di parole tra l'articolo *il* e il sostantivo *mare*). Certo, i risultati sono quantitativamente inferiori rispetto a quelli ricavati secondo i criteri di ricerca impostati in (48), ma danno la certezza di ottenere esclusivamente varianti del proverbio *Tra il dire e il fare c'è di mezzo il mare* scartando a priori qualsiasi frase libera che non abbia a che vedere con il proverbio. Questo è probabilmente il punto focale di questa tappa: riuscire a comprendere le chiavi di ricerca adeguate per escludere il maggior numero di risultati non pertinenti e includere il maggior numero di varianti del proverbio. Vediamo qualche esempio di variante di *Tra il dire e il fare c'è di mezzo il mare* ottenuta con questa ricerca.

- (50) a. Libro sconsigliato agli inguaribili ottimisti dell'oceano internetiano: l'epilogo è emblematico e dimostra come, tra il reale e il virtuale, ci sia davvero **di mezzo il mare**.⁹⁵
- b. Di qui a a ricavarne un progetto architettonico, però, c'è **di mezzo il mare**.⁹⁶

⁹¹ URL: <http://www.cittadinolex.kataweb.it/article_view.jsp?idArt=5383&idCat=128>.

⁹² URL: <<http://web.tiscali.it/cumpustela/giustizia.htm>>.

⁹³ URL: <http://consiglio.regione.emilia-romagna.it/biblio/pagine/num4_2003/Cultura/testo.htm>.

⁹⁴ URL: <http://www.wbt.it/index.php?risorsa=caso_coop>.

⁹⁵ URL: <<http://www.uniurb.it/giornalismo/giornali/febbraio2004/amici.htm>>.

⁹⁶

URL:

<<http://www.cad3d.it/modules.php?op=modload&name=XForum&file=print&fid=37&tid=3202>>. L'esempio non è stato modificato, ecco perché vi si trova una ripetizione (la preposizione *a* ripetuta due volte consecutive dopo *di qui*).

c. Ecco perché avevo aspettato con tanta ansia anche questo 2 ° PD, ecco perché, nonostante tra me e voi "ci sia **di mezzo il mare**", non ho quasi voluto considerare l'ipotesi di potermelo perdere.⁹⁷

d. Perdonare è facile, ma ri-fidarsi ciecamente di qualcuno che ti ha ingannato, lo è meno. Credo che tra perdono e fiducia ci sia **di mezzo il mare!!!!**⁹⁸

Notiamo che vi è una differenza non trascurabile tra gli esempi proposti in (49) e quelli riportati in (50): negli esempi del primo gruppo, *il dire e il fare* sono sinonimi di *parola*, o *discorsi*, e *atti pratici*, oppure di *apprendimento*, o *formazione teorica*, e *pratica*. In altri termini, essi vengono presi alla lettera senza molte implicazioni paremiologiche o senza tanti rimandi al proverbio. Si sente una volontà di ricerca di un qualcosa di idiomatico (*il dire e il fare*, anche per il solo fatto di essere enunciati in questi termini, sono, per così dire, idiomatici), ma non si riesce ad andare oltre alla semplice idiomaticità o alla mera *collocazione*. Al contrario, nel gruppo di proverbi di (50), si osservano importanti modifiche, ma il riferimento al proverbio è assolutamente immediato. Questo viene probabilmente ottenuto grazie alla lontananza semantica tra i primi termini del proverbio (*dire e fare*, nelle sue varie e molteplici declinazioni) e il *mare*, preceduto dall'espressione *di mezzo il*. Il proverbio, noto nella sua variante standard ai nativi italiani o a coloro che hanno una profonda competenza linguistiche in italiano, è implicitamente presente sia per l'emittente sia per il destinatario e funge da chiave di volta che sorregge le nuove frasi inventate (*tra il reale e il virtuale, ci sia davvero di mezzo il mare; da qui a (...), però, c'è di mezzo il mare; tra me e voi "ci sia di mezzo il mare"; tra perdono e fiducia ci sia di mezzo il mare*). Il proverbio, infatti, è l'unico in grado di fornire una spiegazione all'evidente ma apparente incoerenza semantica delle frasi che sono ricalcate su di esso.

È necessario, a questo punto, fare una piccola precisazione. Per rimanere fedeli al nostro proposito iniziale, ovvero permettere ai nostri dati di evidenziare autonomamente le loro regolarità e ricorrenze senza influire direttamente su questo percorso, abbiamo deciso che fosse necessario prendere in considerazione sia il maggior numero di varianti possibile, sia quei casi di creazione e modifica di proverbi. Come già ribadito a più riprese, abbiamo

⁹⁷ URL: <http://www.publinet.it/diabete/forum/pdday_02_2.html>. Il virgolettato era presente nel testo originale.

⁹⁸ URL: <http://www.encanta.it/forum/post.asp?method=ReplyQuote&FORUM_ID=18&CAT_ID=5&Forum_Title=Cuori+solitari&TOPIC_ID=365&REPLY_ID=2704&Topic_Title=Un+classico%3Acome+farsi+perdonare+una+scappatella>.

sempre proceduto senza pregiudizi o preconcetti, facendoci guidare dai dati. Per questa ragione, abbiamo deciso di cercare i proverbi non tanto secondo le varianti comunemente attestate nei dizionari, quanto piuttosto nelle forme effettivamente presenti nei corpora, comprendendo i casi di modifica dei proverbi. Continueremo a chiamare queste frasi, che prendono spunto dai proverbi, con il nome di *proverbio*, a cui aggiungeremo l'aggettivo *modificato*, sebbene siamo consapevoli del filone di letteratura paremiologica che non ritiene che tali frasi siano proverbi.

4.3. Fase di ricerca di proverbi senza l'uso dei marcatori

Dopo aver individuato il campione di proverbi da cercare nei due corpora, nonché il metodo da seguire nella ricerca di tali proverbi, tanto secondo le loro forme standard quanto secondo le loro varianti, ivi comprese le loro versioni modificate, abbiamo proceduto al loro reperimento. Per prima cosa, ci concentreremo sulla lista dei proverbi cercati, con le loro occorrenze iniziali (quelle, cioè, relative al loro utilizzo concomitante con la parola *proverbio* e *proverbe*) e le loro occorrenze totali nel corpus, fornendo anche le chiavi di ogni nostra ricerca. Nelle seguenti due tabelle di Fig. 27 e Fig. 28 sono presenti tutti i proverbi cercati, in itWaC, nella prima tabella, e in frWaC⁹⁹, nella seconda, ordinati per il numero decrescente delle occorrenze totali in itWaC e in frWaC. Notiamo una cosa molto importante: il proverbio più frequente sia in itWaC sia in frWaC è *L'unione fa la forza* (con 509 occorrenze totali) e l'equivalente francese *L'union fait la force* (con 540 occorrenze totali). Si potrebbe trattare solamente di un caso, oppure di una prova di un certo grado di somiglianza tra le due lingue e culture esaminate.

⁹⁹ Si noterà che per la parte relativa a frWaC è stato eliminato un proverbio: *Qui va lentement va sûrement*. Questo perché grazie allo studio delle sue varianti ci siamo resi conto di quanto sia problematico: si tratta, infatti, di uno dei proverbi più frequenti in francese se cercato secondo le diverse varianti, ma tra i meno frequenti se cercato come proverbio. Due di queste varianti sono: *Lentement mais sûrement* e *Doucement mais sûrement*. Tuttavia, sebbene queste varianti abbiano frequenze d'uso assai elevate, esse vengono spesso utilizzate al confine tra l'espressione idiomatica, in quanto costituenti, e il proverbio, in quanto frasi, senza che tale distinzione sia chiara. Tale comportamento problematico ci ha fatto propendere a favore dell'eliminazione del proverbio per la parte relativa alla ricerca dei proverbi nei corpora senza l'uso di marcatori di proverbi.

Proverbi trovati inizialmente con il marcatore <i>proverbio</i>	Occorrenze iniziali	Numero di varianti iniziali	Chiavi di ricerca	Occorrenze totali in itWaC	Numero di varianti in itWaC
l'unione fa la forza	13	1	"unione" "f.*" "la" "forza"	509	1
meglio tardi che mai	5	1	"meglio" "tardi" "che" "mai"	460	1
il buon giorno si vede dal mattino	7	2	il []{1,2} "si" "vede" "dal" "mattino"	366	11
prevenire è meglio che curare	4	2	"prevenire" []{0,4} "che" [tag="Vmn"]	330	55
il fine giustifica i mezzi	4	1	i.* "fin.*" "giustific.*" "i" "mezzi"	310	6
al peggio non c'è fine	4	2	"al" "peggio" "non" []{1,10} "fine" "non" "c" []{0,5} "al" "peggio"	250	19
mal comune, mezzo gaudio	6	2	"mal" "comune"	248	33
l'abito non fa il monaco	8	4	"abito" []{0,1} "fa" "il" "monaco"	223	6
tra il dire e il fare c'è di mezzo il mare	11	4	"di" "mezzo" "il" []{0,3} "mare"	213	30

Proverbi trovati inizialmente con il marcatore <i>proverbio</i>	Occorrenze iniziali	Numero di varianti iniziali	Chiavi di ricerca	Occorrenze totali in itWaC	Numero di varianti in itWaC
dimmi con chi vai e ti dirò chi sei	6	4	"e" "ti" "dir.*" "chi" "sei"	210	108
non tutto il male vien per nuocere	9	4	"non" "tutt.*" "i.*" "mal.*" "v.*" "per" "nuocere"	210	7
il lupo perde il pelo ma non il vizio	9	3	"perde" "il" "pelo" []{1,3} "il" "vizio"	200	23
la mamma dei cretini è sempre incinta	8	5	"de.*" []{0,3} "è" "sempre" "incinta"	200	35
una rondine non fa primavera	6	4	"non" "fa" "primavera"	184	31
se son rose fioriranno	4	2	"se" "son" "rose"	155	9
non si può avere la botte piena e la moglie ubriaca	4	4	"botte" "piena" []{1,10} "ubriaca"	143	6
l'appetito vien mangiando	5	2	[tag="T"] [tag="Nc"] "vien" "mangiando"	141	8
non c'è due senza tre	16	5	"non" "c" "è" [tag="Dn"] "senza" + "non" "c" "è"	125	2

Proverbi trovati inizialmente con il marcatore <i>proverbio</i>	Occorrenze iniziali	Numero di varianti iniziali	Chiavi di ricerca	Occorrenze totali in itWaC	Numero di varianti in itWaC
			[tag="M"] "senza"		
sbagliando s'impara	8	1	"sbagliando" [] {0,1} "impara"	125	3
errare è umano, perseverare è diabolico	7	6	"perseverare" [] {0,5} "√@" "diabolico"	125	14
fatta la legge trovato l'inganno	5	3	"trovato" "l" "inganno"	124	13
non c'è peggior sordo di chi non vuole sentire	7	2	"non" "c" "è" "peggior" "sordo"	120	9
chi si contenta gode	7	3	"chi" "si" ".*contenta" [] {0,3} "gode"	118	5
una mela al giorno leva il medico di turno	11	7	"il" "medico" "di" "turno"	117	50
chi ben comincia è a metà dell'opera	11	6	"chi" "ben" ".*ia"	100	13
morto un papa, se ne fa un altro	7	3	"mort.*" "un.*" [] {1,4} "se" "ne" "fa.*"	94	25
anno nuovo vita nuova	6	3	"anno" "nuovo" [] {1,4} "nuov.*" +	93	32

Proverbi trovati inizialmente con il marcatore <i>proverbio</i>	Occorrenze iniziali	Numero di varianti iniziali	Chiavi di ricerca	Occorrenze totali in itWaC	Numero di varianti in itWaC
			nuov.* "vita" "nuova"		
nessuno è profeta in patria	7	4	"n.*" []{0,3} "profeta" "in" "patri.*"	93	16
chi tace acconsente	5	2	"chi" []{1,5} "acconsente"	92	6
fidarsi è bene, ma non fidarsi è meglio	5	2	"è" "bene" []{0,5} "non" "fidarsi"	89	8
il diavolo fa le pentole ma non i coperchi	4	2	diavolo "fa" "le" "pentole"	84	14
l'occasione fa l'uomo ladro	7	3	"occasion.*" []{0,5} "l" "uomo" "ladro"	80	8
squadra che vince non si cambia	5	2	"squadra" []{0,4} "vinc.*" "non" "si" "cambia"	70	3
il meglio è nemico del bene	5	3	[tag="Nc"] "è" "nemico" "del"	68	6
chi lascia la via vecchia per la nuova, sa quel che lascia e non sa quel che trova	8	7	[tag="T"] [tag="Nc"] "vecch.*" "per" ".*l.*" "nuov.*"	68	9

Proverbi trovati inizialmente con il marcatore <i>proverbio</i>	Occorrenze iniziali	Numero di varianti iniziali	Chiavi di ricerca	Occorrenze totali in itWaC	Numero di varianti in itWaC
a pensar male si fa peccato ma si indovina	4	4	"a" "pensar.*" [] {0,10} "si" "indovina" "a" "pensar.*" [] {0,10} "non" "si" "sbaglia"	64	19
chi non risica, non rosica	5	2	"chi" "non" "risica" "non" "rosica"	61	2
non vendere la pelle dell'orso prima di averlo preso	4	4	"vend.*" "la" "pelle" "dell" "orso" "prima"	61	19
meglio soli che male accompagnati	7	2	"meglio" "soli"	58	9
Natale con i tuoi, a Pasqua con chi vuoi	5	2	".*atale" "con" "i" "tuoi"	53	13
patti chiari, amicizia lunga	4	3	patti [tag="A"] [] {1,3} "amicizia"	52	8
paese che vai usanza che trovi	6	2	"paese" "che" "vai" [] {1,5} "che" "trovi"	49	13
chi fa da sé fa per tre	6	3	"chi" "fa" "da" "s.*" "fa"	46	2
chi rompe paga e i cocci sono suoi	7	3	"chi" "rompe" "paga"	46	6

Proverbi trovati inizialmente con il marcatore <i>proverbio</i>	Occorrenze iniziali	Numero di varianti iniziali	Chiavi di ricerca	Occorrenze totali in itWaC	Numero di varianti in itWaC
chi troppo vuole nulla stringe	6	3	"chi" "troppo" "vuole"	45	8
chi di spada ferisce di spada perisce	4	2	"chi" "di" [tag="Nc"] [tag="Vmi"] "di" [tag="Nc"] "perisce"	42	19
fa più rumore un albero che cade di una foresta che cresce	4	4	"rumore" "un" "albero" "che" "cade"	41	16
di buone intenzioni è lastricato l'inferno	5	5	"buone" "intenzioni" [] {0,6} "inferno"	40	32
chi trova un amico, trova un tesoro	8	2	"chi" "trova" "un.*" [] {1,4} "trova"	40	4
aiutati che Dio t'aiuta	13	4	"aiutati" "che" [] {0,5} "aiuta"	38	4
due x sono meglio di uno	6	3	"due" [] {1,3} "sono" "meglio" "di"	34	28
l'ozio è il padre dei vizi	4	3	"ozio" [] {0,5} "padre" [] {0,5} "vizi"	34	11
moglie e buoi dei paesi tuoi	6	2	"mogli.*" "e" "buoi" "dei"	33	1

Proverbi trovati inizialmente con il marcatore <i>proverbio</i>	Occorrenze iniziali	Numero di varianti iniziali	Chiavi di ricerca	Occorrenze totali in itWaC	Numero di varianti in itWaC
			"paesi" "tuoi"		
se la montagna non va da Maometto, Maometto va alla montagna	6	4	"se" ".*aometto" "non" "va" + "se" "la" "montagna" "non" "va"	30	15
ride bene chi ride ultimo	4	2	"ride" "bene" "chi" "ride"	30	2
chi dorme non piglia pesci	5	2	"chi" "dorme" "non" "piglia" "pesci"	29	1
la gatta frettolosa fece i gattini ciechi	8	5	"la" "gatta" "frettolosa"	29	12
tanto va la gatta al lardo che ci lascia lo zampino	4	3	"tanto" "va" "la" "gatta" "a.*" "lar.*o"	28	6
gallina vecchia fa buon brodo	5	2	"gallina" "vecchia" "fa"	27	3
a caval donato non si guarda in bocca	5	1	"a" "caval" "donato" "non" "si" "guarda"	27	1
can che abbaia non morde	5	1	"can" "che" "abbaia" "non" "morde"	26	1
chi va piano va sano e	10	3	"chi" "va"	25	5

Proverbi trovati inizialmente con il marcatore <i>proverbio</i>	Occorrenze iniziali	Numero di varianti iniziali	Chiavi di ricerca	Occorrenze totali in itWaC	Numero di varianti in itWaC
va lontano			"piano" "va"		
chi più spende, meno spende	4	2	"chi" "più" "spende" "meno"	24	2
sposa bagnata sposa fortunata	5	4	"spos.*" "fortunat.*"	23	5
quando il gatto non c'è i topi ballano	5	4	"quando" "il" "gatto" "non"	22	4
occhio non vede cuore non duole	7	3	"occhio" "non" "vede" "cuore" "non"	22	1
tra moglie e marito non mettere il dito	4	4	".*ra" [tag="Nc"] "e" [tag="Nc"] "non" []{0,1} "mettere"	22	10
chi pecora si fa, il lupo se la mangia	4	4	"pecora" []{1,6} "lupo" []{1,6} "mang.*"	21	9
impara l'arte e mettila da parte	4	2	"impar.*" "l" "arte" "e" "mett.*" "da" "parte"	17	3
dagli amici mi guardi iddio, ché dai nemici	5	5	".*mici" []{0,5} ".*io" []{0,5}	17	7

Proverbi trovati inizialmente con il marcatore <i>proverbio</i>	Occorrenze iniziali	Numero di varianti iniziali	Chiavi di ricerca	Occorrenze totali in itWaC	Numero di varianti in itWaC
mi guardo io			".*mici" "mi" "guard.*"		
non svegliare il cane che dorme	4	4	"non" [tag="Vmn"] [tag="T"] [tag="Nc"] "che" "dorme"	16	10

Fig. 27. Lista di tutti i proverbi cercati in itWaC e ordinati secondo le relative occorrenze totali. Nella prima colonna troviamo tutti i proverbi, nella seconda colonna le occorrenze iniziali di ognuno di questi, ovvero il numero di volte in cui essi comparivano assieme alla parola *proverbio*, e nella terza le loro varianti iniziali. Nella quarta colonna si offrono le chiavi di ricerca utilizzate per cercare i proverbi, mentre nella quinta le occorrenze totali in itWaC, per terminare con la sesta colonna in cui si annota il numero di varianti di ogni proverbio presenti nell'intero corpus.

Proverbi trovati inizialmente con il marcatore <i>proverbio</i>	Occorrenze iniziali	Numero di varianti iniziali	Chiavi di ricerca	Occorrenze totali in frWaC	Numero di varianti in frWaC
l'union fait la force	11	2	"union" "fait" "la" "force"	540	1
avoir le beurre l'argent du beurre et la crème	4	3	"beurre" []{0,3} "argent" "du" "beurre"	309	24
l'habit ne fait pas le moine	11	3	habit []{1,6} "le" "moine"	274	41
mieux vaut tard que jamais	7	2	"mieux" "vaut" "tard" "que" "jamais"	274	1

Proverbi trovati inizialmente con il marcatore proverbio	Occorrenze iniziali	Numero di varianti iniziali	Chiavi di ricerca	Occorrenze totali in frWaC	Numero di varianti in frWaC
il n'y a pas de fumée sans feu	8	5	fum.* "sans" "feu"	202	23
il vaut mieux prévenir que guérir	10	6	"mieux" []{0,3} "pr√©venir" "que" "gu√©rir"	201	5
tout vient à point à qui sait attendre	10	4	à "qui" "sait" "attendre"	180	18
on n'est jamais si bien servi que par soi même	5	3	"est" "jamais" []{1,6} "servi.*" "que" "par" "soi"	169	4
nul n'est prophète en son pays	8	3	prophète []{1} "son" "pays"	163	32
le temps c'est de l'argent	5	2	"le" "temps" []{1,5} "de" []{1,5} "argent"	158	14
l'argent ne fait pas le bonheur	8	3	argent "ne" "fait" "pas" "le" "bonheur"	150	11
jamais deux sans trois	7	3	"jamais" [tag="M"] "sans" [tag="M"]	144	5
il ne faut pas vendre la peau de l'ours avant de l'avoir tué	5	5	"pas" []{0,3} "l.*" "peau" "de" []{0,3} "ours"	140	24

Proverbi trovati inizialmente con il marcatore proverbio	Occorrenze iniziali	Numero di varianti iniziali	Chiavi di ricerca	Occorrenze totali in frWaC	Numero di varianti in frWaC
qui ne tente rien n'a rien	10	3	qui "n.*" [] {1,3} "rien" "n" "a" "rien"	128	6
c'est en forgeant qu'on devient forgeron	12	3	"en" "forgeant" [] {0,3} "devient"	104	8
plus on est de fous, plus on rit	4	4	"plus" "on" "est" "de" "fous"	101	21
l'enfer est pavé de bonnes intentions	4	2	enfer "est" "pavé" "de"	98	5
les petits ruisseaux font les grandes rivières	4	2	"petit.*" "ruisseau.*" "f.*" "l.*" "grand.*" "rivi@r.*"	95	5
quand on aime on ne compte pas	7	2	"quand" "on" "aime" "on" "ne" "compte"	89	2
en avril ne te découvre pas d'un fil	9	6	"avril" [] {0,2} "ne" "te" "découvre.*" "pas" [] {1,2}	85	3
tous les chemins mènent à rome	4	1	"tous" "les" "chemins" "mènent"	82	37
qui veut noyer son chien l'accuse de la rage	12	9	veut [] {1,4} "son" "chien" [] {1,6} "rage"	69	16

Proverbi trovati inizialmente con il marcatore proverbio	Occorrenze iniziali	Numero di varianti iniziali	Chiavi di ricerca	Occorrenze totali in frWaC	Numero di varianti in frWaC
quand le sage montre la lune l'idiot regarde le doigt	4	3	"montre" "la" "lune" []{0,6} "regarde" "le" "doigt"	66	22
charité bien ordonnée commence par soi-même	8	4	charité "bien" "or.*ée"	65	6
qui aime bien châtie bien	8	4	qui "aime" "bien" []{1} "bien"	62	1
loin des yeux, loin du cœur	14	3	loin "des" "yeux" []{1,5} "du" "c.*ur"	56	10
qui se ressemble s'assemble	9	1	qui "se" "ressemble" []{0,3} "assemble"	52	1
les cordonniers sont toujours les plus mal chaussés	5	3	"cordonnier.*" []{0,4} "le.*" "plus" "mal" "chauss√©.*"	51	17
qui ne dit mot consent	4	2	"qui" "ne" "dit" "mot" "consent"	50	1
la nuit porte conseil	6	1	"la" "nuit" "porte" "conseil"	43	1
la critique est aisée, et l'art est difficile	4	4	"est" "aisé.*" []{1,6} "est"	43	5

Proverbi trovati inizialmente con il marcatore proverbio	Occorrenze iniziali	Numero di varianti iniziali	Chiavi di ricerca	Occorrenze totali in frWaC	Numero di varianti in frWaC
			"difficile"		
c'est dans les vieux pots qu'on trouve les bonnes soupes	5	4	"est" "dans" "les" [tag="A"] [tag="Nc"] [] {1,3} "fait" "les"	40	6
aide-toi le ciel t'aidera	7	5	"aide.*" [] {0,1} ".*toi" [] {1,20} "aid.*"	37	8
à chacun son métier et les vaches seront bien gardées	4	4	"et" "les" "vaches" "seront" "bien" "gard.*"	31	20
les chiens aboient, la caravane passe (Maloux)	4	4	"les" [] {1,7} "l.*" "caravane.*" "passe.*"	29	9
qui sème le vent récolte la tempête	7	2	"qui" "sème" "le" "vent"	26	4
divise afin de régner (Maloux)	4	2	divis.* "pour" [] {0,2} "regner"	23	2
tel père tel fils	6	3	"tel" [tag="Nc"] "tel" [tag="Nc"]	22	11
tant va la cruche à l'eau qu'à la fin elle se casse	6	2	"tant" "va" "la" "cruche"	22	3
qui trop embrasse mal étreint	6	2	"qui" "trop" "embras.*e"	17	1

Proverbi trovati inizialmente con il marcatore proverbio	Occorrenze iniziali	Numero di varianti iniziali	Chiavi di ricerca	Occorrenze totali in frWaC	Numero di varianti in frWaC
			"mal"		
qui vole un œuf vole un bœuf	4	2	qui "vole" "un" [] {1,4} "vole" "un"	14	2
dis-moi ce que tu manges, et je te dirai ce que tu es	4	4	"di.*" [] {1,3} ".*moi" [] {0,9} "je" [] {1,3} "dirai"	13	13
qui dort dîne	4	2	"qui" "dort" "d.*ne"	10	1
en mai fais ce qu'il te plait	6	4	"en" "mai" [] {0,3} "fais" [] {0,3} "ce" [] {0,6} "pla.*"	9	1
la peur donne des ailes	4	1	la "peur" "donne"	6	2

Fig. 28. Lista di tutti i proverbi cercati in frWaC e ordinati secondo le relative occorrenze totali. Nella prima colonna troviamo tutti i proverbi, nella seconda colonna le occorrenze iniziali di ognuno di questi, ovvero il numero di volte in cui essi comparivano assieme alla parola *proverbe*, e nella terza le loro varianti iniziali. Nella quarta colonna si offrono le chiavi di ricerca utilizzate per cercare i proverbi, mentre nella quinta le occorrenze totali in frWaC, per terminare con la sesta colonna in cui si annota il numero di varianti di ogni proverbio presenti nell'intero corpus.

Analizziamo le due tabelle presentate per considerare alcuni valori importanti di grande interesse non solo per questo studio ma anche, in generale, per le future ricerche che auspichiamo vengano eseguite in tale ambito. *In primis*, balza agli occhi un dato: l'occorrenza iniziale dei proverbi, quella cioè determinata dall'uso concomitante dei marcatori di proverbio, non rispecchia le occorrenze totali dei proverbi. Prendiamo un dato valido per tutti: il proverbio *Aiutati che Dio t'aiuta*, con frequenza iniziale pari a 13, si trova verso il fondo della

graduatoria delle occorrenze reali in itWaC, con solo 38 occorrenze, contro le 509 occorrenze dell'altro proverbio con frequenza iniziale 13, ossia *L'unione fa la forza*, che si aggiudica il primo posto nella nostra classifica delle occorrenze reali. Per maggiore chiarezza, abbiamo deciso di presentare esclusivamente i dati numerici nella prossima tabella (Fig. 29).

Occorrenze iniziali	4	5	6	7	8	9	10	11	13	16
Occorrenze reali riscontrate in itWaC	330	460	248	366	223	210	25	213	509	125
	310	141	210	125	200	200		117	38	
	250	124	184	120	125			100		
	155	92	93	118	68					
	143	89	49	94	40					
	84	70	46	93	29					
	64	68	45	80						
	61	61	34	58						
	52	53	33	46						
	42	40	30	22						
	41	29								
	34	27								
	30	27								
	28	26								
	24	23								
	22	22								
	21	17								
17										
16										

Fig. 29. Raffronto tra le occorrenze iniziali dei proverbi, individuate cioè con la parola *proverbio*, e la frequenza reale, effettuando le ricerche già illustrate nell'intero corpus itWaC.

Per maggiore chiarezza, presentiamo nella prossima tabella (Fig. 30) la frequenza media effettiva per ogni occorrenza iniziale.

Occorrenza iniziale (itWaC)	Media occorrenze reali (itWaC)
4	91
5	81
6	97

7	112
8	114
9	205
10	25
11	143
13	274
16	125

Fig. 30. Tabella relativa alla media delle frequenze reali trovate per ogni occorrenza iniziale.

Da questi dati emerge la non corrispondenza tra le occorrenze dei proverbi introdotti dalla parola *proverbio* e le occorrenze effettive dei proverbi nel corpus. Pertanto, sebbene le prime occorrenze trovate grazie al marcatore di proverbio possano essere indicative di una minore o maggiore probabilità di reperire una quantità minima di proverbi (da noi individuata su una soglia di 20 occorrenze) in funzione di una minore o maggiore occorrenza iniziale, non possiamo ritenere che queste informazioni iniziali siano altrettanto indicative dell'andamento quantitativo dei relativi proverbi nell'intero corpus di riferimento. Questo fa capire quanto la fase iniziale di reperimento dei proverbi sia di grande utilità per individuare il campione di proverbi da ricercare nel corpus, ma non si possa rivelare altrettanto indicativa del reale comportamento dei proverbi non accompagnati dalla parola *proverbio*.

Questi dati non fanno che sottolineare un'altra questione di estrema importanza per il funzionamento dei proverbi nel discorso: essi sono raramente accompagnati da un qualche marcatore che ne segnala lo status paremiologico, il che conferma quanto già esposto da Kleiber (1999), Shapira (2000) e Cram (1983).¹⁰⁰

4.4. Altri marcatori di proverbio

Abbiamo quindi cercato qualche altro marcatore di proverbio che potesse essere utilizzato in studi futuri per reperire automaticamente il maggior numero di proverbi possibile e realizzare una ricerca simile alla nostra, ma che consenta di giungere a migliori risultati quantitativi e qualitativi.

Alla luce di quanto espresso in questo e negli scorsi capitoli, nonché sulla base dei dati reperiti in itWaC nella seconda fase di ricerca, crediamo che la nostra scelta iniziale sia stata mirata, sebbene ammettiamo che il procedimento risulti ancora troppo macchinoso e generi una quantità ancora elevata di scarti. Gli scarti a cui ci riferiamo non sono solamente quelli

¹⁰⁰ Si veda anche § 1.6.

derivanti dall'impossibilità di prendere in considerazione tutti gli esempi ottenuti, in quanto facenti parte dei casi che abbiamo eliminato nella fase di scrematura delle concordanze della parola *proverbio*, bensì anche tutti quei proverbi (che come abbiamo visto sono la maggior parte) che hanno solamente una frequenza e che non garantiscono di essere veri proverbi o, quanto meno, di essere riconosciuti come tali dall'intera comunità linguistica.

Per cercare ulteriori esempi di marcatori di proverbio, abbiamo creato un sottocorpus contenente i risultati della ricerca dei proverbi nell'intero corpus. In questo modo, abbiamo potuto riunire i dati relativi ai proverbi nelle loro versioni standard e in quelle modificate.

Escludendo i risultati che contengono la parola *proverbio*, osserviamo qualche ricorrenza importante: i proverbi sono spesso introdotti da espressioni quali *Come si dice*, *Come dire*, *Come si usa dire*, *Come si suol dire*, *Come si sa*, ma anche *Come sapete* o *Come al solito*. Al posto della parola *proverbio* si nota un uso frequente della parola *detto*, usata, quindi, per indicare non solo i detti, ma anche i proverbi; lo stesso vale per la parola *espressione* che viene utilizzata in funzione di marcatore di proverbio. In realtà, l'idea di utilizzare altre etichette per segnalare l'enunciazione di un proverbio ci era già stata indicata dal prof. Wolfgang Mieder, insigne paremiologo di fama mondiale, tra l'altro fondatore della rivista *Proverbium*. In un colloquio privato con il professore nell'ambito della Conferenza Internazionale sui Proverbi, nel 2015, a Tavira (Portogallo), è emerso, da parte del Professore, il suggerimento di cercare i proverbi non solamente sulla base della parola *proverbio* ma anche utilizzando altre etichette, come, ad esempio, quelle appena mostrate (*detto* ed *espressione*). Certo, il consiglio è sicuramente ottimo e molto interessante, in quanto è un dato di fatto che quando il parlante medio enuncia proverbi, in realtà li indica utilizzando altre categorie. Ciononostante, abbiamo rifiutato l'utilizzo di queste etichette, ma anche dei marcatori mostrati in questo paragrafo, in quanto generano ancora più errori, confusione e scarti rispetto alla parola *proverbio*. Se, infatti, inseriamo le stringhe *Come si suol dire*, *Come si sa*, ecc., oppure semplicemente *il detto* o ancora *l'espressione*, otteniamo senz'altro proverbi ma anche tantissimi altri risultati da analizzare, valutare e probabilmente eliminare. Con il marcatore *Come si suol dire*, ad esempio, si trovano tantissime espressioni idiomatiche, che verrebbero quindi eliminate in un'analisi, come la nostra, tutta relativa ai proverbi.

Certamente questo punto dovrà essere sviluppato anche alla luce di tutte le nostre considerazioni. Per ora, possiamo comunque affermare che la nostra scelta iniziale viene confermata da questi ulteriori dati ottenuti nella seconda fase del nostro studio.

4.5. Analisi del nostro marcatore di proverbio e riflessioni sulle risorse dedicate ai proverbi

Capire che cosa si indichi comunemente per proverbio è sicuramente di grande interesse: durante la fase di scrematura delle concordanze della parola *proverbio* e *proverbe*, e nella successiva composizione della nostra banca dati di proverbi, abbiamo riscontrato diversi casi, tanto in italiano quanto in francese, in cui si indicano, con la parola proverbio, elementi che non possono essere considerati tali in ambito linguistico o paremiologico. Tuttavia, non essendo questo uno dei nostri obiettivi, non abbiamo compiuto una vera e propria ricerca quantitativa su questo aspetto: si tratta, dunque, di un ambito di ricerca che merita ulteriori approfondimenti.

Ad ogni modo, è possibile operare su questo punto almeno due riflessioni diverse: da un lato, i parlanti mostrano, in varie occasioni, una scarsa competenza paremiologica, dall'altro sono i dizionari di proverbi che non sempre stanno al passo con i tempi.

Accade spesso, ad esempio, che gli autori dei campioni di testi da noi analizzati, indichino, con la parola proverbio, non tanto frasi quanto piuttosto costituenti di frasi. Vediamo qualche esempio:

- (51) (...) e il nostro proverbio del «Cadere dalla padella nella brace» diventa, in arabo, «Kal mustajir minar Ramadaa binnar». ¹⁰¹
- (52) Con milioni di pagine presenti oggi sulla Rete, in molti casi è, come dice il proverbio, come cercare il famoso ago nel pagliaio. ¹⁰²
- (53) Tu sai mio caro, il nostro vecchio proverbio; 'restare in piedi anche nel fango;' applichamocelo, sfidiamo i rimproveri dei fratelli per servire la causa comune. ¹⁰³
- (54) Anche le noci non hanno scherzato, sono state segnalate a 5.40 euro al chilo e i fichi secchi a 7 euro. Altroché "vale un fico secco", bisognerà cambiare il proverbio. ¹⁰⁴
- (55) De là, ce proverbe : être désœuvré comme Ali Ben Hassan. ¹⁰⁵
- (56) (...) dont on espère qu'ils vont faire parler le proverbe "ne jamais négliger le coeur d'un champion". ¹⁰⁶

¹⁰¹ URL: <<http://www.01-4.eurnews.it/ottobre/lastampa19.htm>>.

¹⁰² URL: <<http://www.comune.siena.it/bandini/FBI6.html>>.

¹⁰³ URL: <<http://www.braidense.it/scaffale/genio4.html>>.

¹⁰⁴ URL: <<http://www.codacons.it/articolo.asp?id=7191>>.

¹⁰⁵ URL: <<http://www.jose-corti.fr/titresromantiques/histoire-du-prince-ahmed.html>>.

(57) Il se plaisait à demander qu'on laissât du temps au temps.¹⁰⁷

Negli esempi da (51) a (57), si definisce proverbio ciò che sicuramente non può essere tale: *Cadere dalla padella nella brace*, *Cercare il famoso ago nel pagliaio*, *Restare in piedi anche nel fango*, *Être désœuvré comme Ali Ben Hassan*, *Ne jamais négliger le cœur d'un champion* e *Demander qu'on laisse du temps au temps* sono tutti costituenti e non frasi. Per divenire frasi, ed essere enunciati, necessitano almeno di un soggetto, come succede negli esempi riportati.

Oltre a questi casi, abbiamo riscontrato esempi in cui si indica l'enunciatore primo dei proverbi, cosa che è in forte contraddizione con una delle prerogative dei proverbi, quella già trattata in § 1.4.1, secondo la quale i proverbi sono frasi *ON-sentencieuses* (per usare la terminologia di Anscombe, 2000). Nella pratica, invece, vediamo che molti parlanti indicano i gli autori primi (reali o presunti) dei proverbi, come negli esempi seguenti:

(58) Sfogliando un vecchio giornale mi sono imbattuto su un classico "proverbio" di Frank Zappa che avevo dimenticato, che recita: "La critica rock è gente che non sa scrivere che intervista gente che non ha nulla da dire, per gente che non sa leggere".¹⁰⁸

(59) (...) comme le dit mon proverbe : " quand on tue le coq, les poules s'entretuent. "¹⁰⁹

(60) En effet Carla ne va pas avec lui en Inde, certains signes montrent même que rien n'est consommé avant le mariage, d'où le proverbe de Frédéric DARD il vaut mieux avoir les bourses en action que les actions en bourse.¹¹⁰

(61) (...) ça me rappelle le fameux proverbe, de je ne sais plus trop qui, moi peut-être : " Plus on pédale moins vite, moins on avance plus davantage ".¹¹¹

In questi esempi, che da una rapida analisi sembrano poco numerosi, si osserva la presenza esplicita dell'autore primo dei proverbi citati, caratteristica propria, ad esempio, dell'aforisma. Una prova di ciò sta nel fatto che questi esempi ricorrono solo una volta con la

¹⁰⁶ URL: <http://mondial2006.blogs.liberation.fr/mon_weblog/2006/06/allez_un_petit_.html>.

¹⁰⁷ URL: <<http://generisfrance.free.fr/esgeaa/benoit0598.html>>.

¹⁰⁸ URL: <<http://blog.repubblica.it/rblog/page/EAssante/20031129>>. Il refuso (*intevista*) è presente nel testo originale.

¹⁰⁹ URL: <<http://kohlanta.tf1.fr/kohlanta/forum/0,,3478042-e01FU19JRCBVTIZfSUR9IHsxNjMyOTM5IDF9,00-forum-koh-lanta-.html>>.

¹¹⁰ URL: <<http://blog.repubblica.it/rblog/page/EAssante/20031129>>.

¹¹¹ URL: <<http://richardgre.blog.lemonde.fr/2008/02/05/carnet-de-semaine-choc/>>.

parola *proverbio* e che, in generale, hanno poche occorrenze nei due corpora da noi usati, cosa che indica la probabile consapevolezza dell'intera comunità linguistica del fatto che essi non siano proverbi.

Oltre a ciò, si hanno episodi diametralmente opposti: esistono nei nostri corpora proverbi con molte occorrenze che, però, non vengono considerati affatto dai dizionari. I più significativi di questi sono i seguenti (a fianco a ciascun proverbio si indicano le occorrenze totali in itWaC): *Fa più rumore un albero che cade di una foresta che cresce* (41); *La mamma dei cretini è sempre incinta* (200); *Prevenire è meglio che curare* (330). Per la parte relativa al francese si trovano ancora più esempi di proverbi frequenti in frWaC e non attestati nei principali dizionari di proverbi, il che indica probabilmente il minor grado di completezza delle risorse paremiologiche francesi rispetto a quelle italiane (tra parentesi le occorrenze totali in frWaC): *Tel père tel fils* (22); *Quand le sage montre la lune, l'idiot regarde le doigt* (66); *Quand on aime on ne compte pas* (89); *Qui ne tente rien n'a rien* (128); *Le temps c'est de l'argent* (158); *Qui va lentement va sûrement* (più di 800).

Generalmente, che cosa può significare la mancanza di attestazione di proverbi, reperiti in corpora di riferimento, da parte dei principali dizionari di proverbi? È difficile dare una risposta unica e certa, ma si possono avanzare alcune ipotesi. In primo luogo, si potrebbe sostenere che i proverbi che non trovano attestazione nei dizionari non sono considerati tali dai paremiologi, il che è plausibile vista la non perfetta competenza paremiologica dei parlanti. In secondo luogo, si potrebbe trattare di proverbi che non appartengono a quella lingua e cultura oppure che sono stati inventati. Qualunque sia la risposta, bisognerebbe indagare più a fondo in questo ambito, in quanto alcuni di questi proverbi (si vedano quelli sopra riportati) sono molto frequenti e familiari, come mostrano i dati evidenziati. Ancora una volta, si ribadisce il ruolo fondamentale, per l'intera area della paremiologia, dello studio dei proverbi su dati autentici e reali.

4.6. Proverbi e variabilità

Una delle principali tappe raggiunte grazie alle diverse fasi analisi affrontate in questo studio riguarda la variabilità dei proverbi. Dopo aver notato, già nella prima fase di ricerca, una notevole presenza di proverbi modificati, abbiamo deciso di esaminare da vicino questo fenomeno. Per questa ragione, una volta reperiti i proverbi grazie alle concordanze dei marcatori di proverbio, abbiamo deciso di quantificare i proverbi modificati: questi sono stati calcolati sia sui valori iniziali (ossia, sulle prime occorrenze relative alla combinazione degli

stessi con le parole *proverbio* e *proverbe*), sia sulle occorrenze totali della seconda fase di ricerca. I risultati di questo calcolo sono già stati riportati in Fig. 27 (relativa ai proverbi italiani) e **Errore. L'origine riferimento non è stata trovata.** (riguardante quelli francesi).

La prima domanda che possiamo formulare è la seguente: qual è la relazione tra il grado di variabilità, ovvero la capacità di ogni proverbio di generare varianti e versioni modificate, dei proverbi reperiti con i relativi marcatori di proverbio, rispetto al grado di variabilità degli stessi cercati nei due corpora senza i marcatori di proverbio? Questa risposta può dare una conferma del livello di somiglianza tra le informazioni ottenute nella prima fase dello studio e quelle riscontrate nella seconda fase.

Per rendere più omogenei e comparabili i nostri calcoli, abbiamo coniato l'espressione *indice di variabilità*, che corrisponde al rapporto percentuale tra il numero di varianti (considerando sempre i valori univoci, ossia escludendo le ripetizioni) e il numero di occorrenze. Tale indice sarà, pertanto, direttamente proporzionale al numero delle varianti rispetto alle occorrenze: ciò significa che più varianti per occorrenze saranno attestate e più l'indice di variabilità sarà elevato. Un indice di variabilità tendente allo 0% sarà tipico di quei proverbi che mostrano un numero molto basso di varianti rispetto alle occorrenze totali; al contrario, più tale indice tende al 100% e più varianti rispetto alle occorrenze verranno riscontrate.

Precisiamo, tuttavia, che il valore dello 0% non verrà mai rappresentato, contrariamente al 100% che, invece, trova parecchi riscontri. Questo è dovuto al fatto che nei casi in cui abbiamo reperito una sola variante, che quindi potrebbe essere considerata non una variante ma il proverbio nella sua versione standard, abbiamo indicato quest'unica variante con il valore 1 e non con il valore 0. Secondo alcuni, trattandosi non di una variante ma della forma standard del proverbio, sarebbe forse più logico indicarla con il numero 0. La ragione per cui abbiamo scelto di indicare anche l'unica variante di un proverbio col valore 1 sta nella consapevolezza dei limiti del corpus e della nostra ricerca: basterebbe, infatti, ampliare i criteri di ricerca per trovare ulteriori varianti, cosa che andrebbe a modificare gli indici di variabilità. Ecco perché abbiamo comunque preferito indicare tutte le informazioni certe in nostro possesso. Ad esempio, l'informazione certa è che anche il proverbio più frequente all'interno del corpus itWaC, ossia *L'unione fa la forza*, ha un totale di occorrenze di 509 unità e una sola variante (che è, appunto, *L'unione fa la forza*). L'indice di variabilità verrà così calcolato dividendo il totale delle varianti (1) per il numero di occorrenze totali (509). Il

calcolo è il seguente: $1/509 = 0,00196$; in valore percentuale equivale allo 0,196%, ovvero, arrotondando per eccesso, allo 0,2%. In questa maniera, lasciamo spazio a un eventuale ricalcolo nel caso in cui cambiassero alcune delle variabili da noi utilizzate, come il corpus, ma soprattutto i criteri di ricerca. Proseguiamo l'esempio per dimostrare e chiarire quanto stiamo affermando.

Come indicato in § 4.2, abbiamo proceduto alla ricerca dei proverbi sulla base della o delle varianti emerse nella prima fase di ricerca, quella relativa alla concordanza dei marcatori di proverbio. Il proverbio *L'unione fa la forza* mostrava, nella prima ricerca *corpus-driven*, 13 occorrenze e una sola variante. Pertanto, non abbiamo fatto altro che cercare il proverbio stesso tale e quale, senza dover cercare anche la parola *proverbio* e senza andare a modificare i parametri di ricerca. Questa scelta è stata compiuta per coerenza nei confronti delle regole individuate e seguite per cercare i proverbi all'interno dei due corpora.

Tuttavia, se si compissero ulteriori analisi oppure se si modificassero i parametri di ricerca, andando a cercare per ogni proverbio tutte le possibili varianti (ammesso che ciò sia possibile), si potrebbero reperire altre varianti. Se, infatti, cercassimo il proverbio in oggetto inserendo nella maschera di ricerca la stringa *"unione" "fa" [tag="T"] [tag="Nc"]* oppure *[tag="T"] [tag="Nc"] "fa" "la" "forza"*, dove *[tag="T"]* sta per qualsiasi parola etichettata come articolo determinativo e *[tag="Nc"]* sta per qualsiasi parola *taggata* come nome comune, troveremmo altri risultati relativi al proverbio modificato, come *L'unione fa l'innovazione*, *L'unione fa la cooperativa*, *L'unione fa la farsa*, *L'unione fa la qualità*, *L'unione fa la sicurezza*, *L'unione fa lo sforzo*, oppure *Il gruppo fa la forza*, *Il coordinamento fa la forza*, *L'aggregazione fa la forza*, *L'euro fa la forza*, *L'insieme fa la forza*, *La moltitudine fa la forza*, *La Rete fa la forza*, *La riunione fa la forza*, *L'unità fa la forza*. Il numero delle occorrenze totali aumenterebbe, anche se di poco: anziché essere 509 diventerebbe 544. Il numero delle varianti salirebbe parecchio, in percentuale, poiché anziché essere 1 sarebbe 26. Il nuovo indice di variabilità sarebbe pertanto più elevato e si attesterebbe al 4,8%. Sicuramente molto più elevato rispetto allo 0,2% iniziale, ma ancora abbastanza contenuto, dato il numero elevato di occorrenze.

La consapevolezza di quanto i valori delle occorrenze e delle loro variabili possa cambiare modificando anche solo i parametri di ricerca impone di evitare di indicare lo 0% come indice di variabilità minimo, in quanto troppo facilmente confutabile.

In Fig. 31 e Fig. 32 sono riportati i risultati di quest'ulteriore analisi: il primo grafico si riferisce al confronto tra l'indice di variabilità iniziale e quella totale calcolata sull'intero corpus itWaC e il secondo grafico esprime lo stesso confronto ma sui dati reperiti grazie all'analisi di frWaC. Nell'asse verticale sono riportati gli scaglioni percentuali dell'indice di variabilità e su quello orizzontale sono indicate le diverse occorrenze iniziali. Ciò che emerge chiaramente da entrambi i grafici è che l'andamento dell'indice di variabilità iniziale non segue quello dello stesso indice calcolato sui dati completi reperiti nei due corpora. Un'ulteriore prova di quanto i marcatori di proverbi siano utili per reperire automaticamente proverbi effettivamente in uso in una lingua ma si rivelino meno validi per fornire un'analisi accurata, specialmente quantitativa, degli stessi.

Confronto tra l'indice di variabilità iniziale e quello totale in itWaC

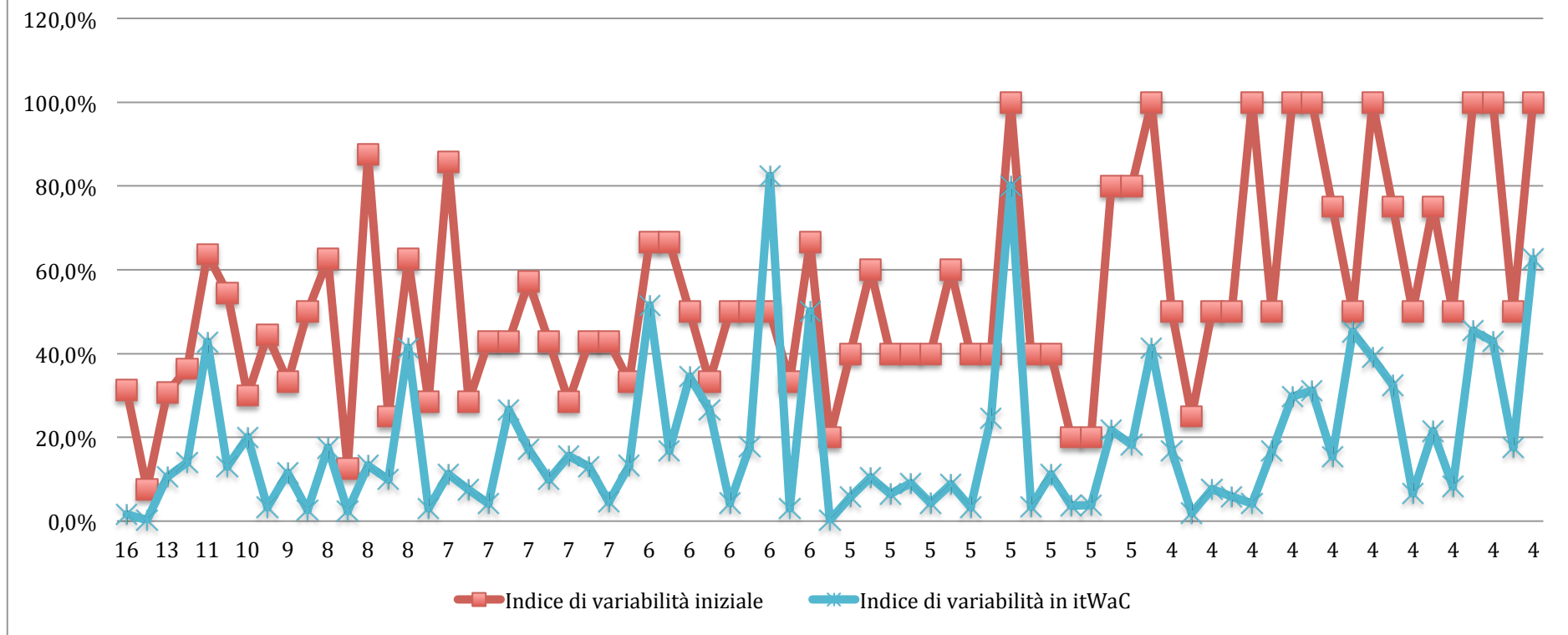


Fig. 31. Confronto tra l'indice di variabilità iniziale (ossia il rapporto tra varianti trovate e il totale di occorrenze dei proverbi introdotti dalla parola proverbio) e l'indice di variabilità totale in itWaC, ottenuto dividendo il numero di varianti reperite per le occorrenze reali nel corpus dei proverbi cercati senza la parola proverbio.

Confronto tra l'indice di variabilità iniziale e quello totale in frWaC

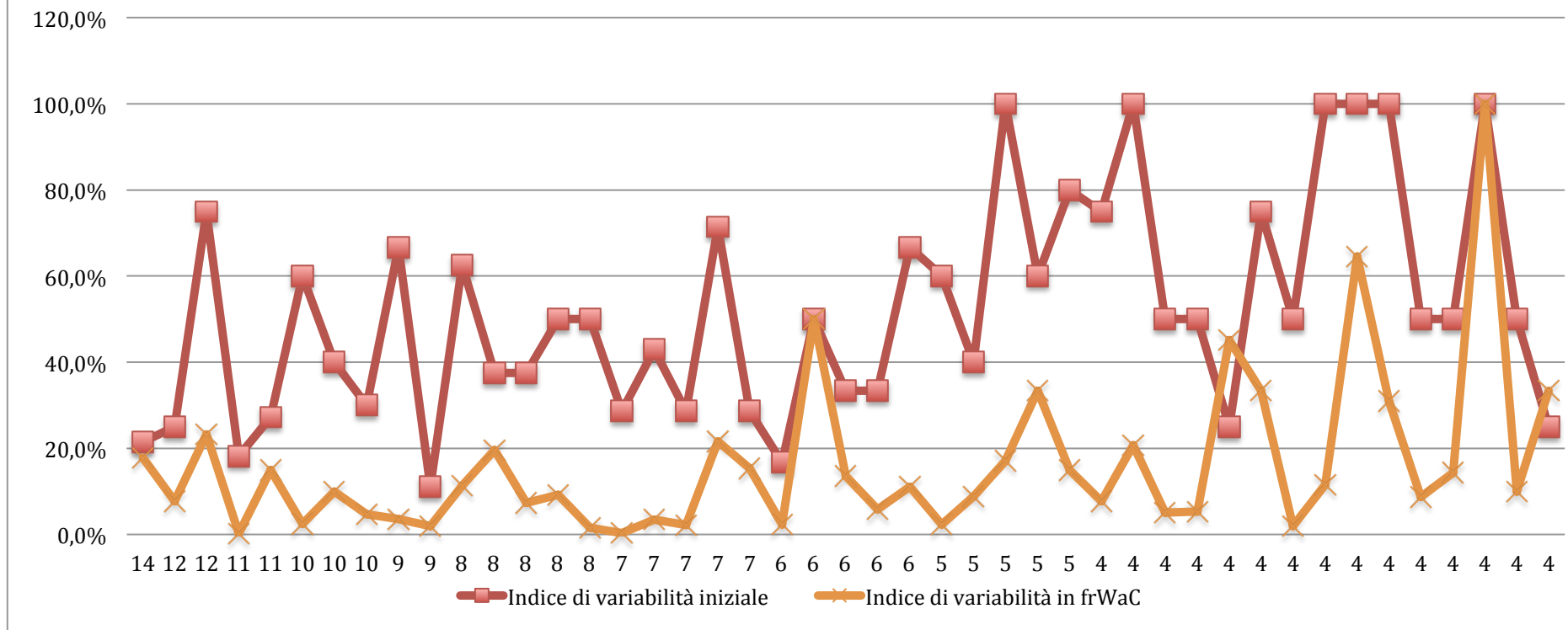


Fig. 32. Confronto tra l'indice di variabilità iniziale (ossia il rapporto tra le varianti trovate e il totale di occorrenze dei proverbi introdotti dalla parola proverbio) e l'indice di variabilità totale in frWaC, ottenuto dividendo il numero di varianti reperite per le occorrenze reali nel corpus dei proverbi cercati senza la parola proverbio.

Una volta abbandonato l'indice di variabilità dei proverbi legato alla loro frequenza iniziale, abbiamo deciso di analizzare comunque questo dato che ci sembrava di grande importanza per fornire un'analisi accurata dei proverbi, sebbene, come già detto, il fenomeno della *variabilità* sia molto frequente e incida non poco sull'uso, sul funzionamento e sull'analisi dei proverbi. Per questa ragione, si è deciso di continuare a lavorare con l'indice di variabilità ma solamente relativo alle occorrenze totali dei proverbi reperiti nei due corpora.

Ci siamo, così, chiesti se ci fosse una relazione tra la frequenza d'uso dei proverbi e la loro capacità di generare varianti. Prima di procedere all'analisi vera e propria, sono state fatte delle previsioni in funzione di nostri ragionamenti logici e sulla base di quanto appreso in ambito paremiologico dalla letteratura esistente.

Per prima cosa, si è pensato che a una frequenza d'uso maggiore corrispondesse un indice di variabilità maggiore. Questo perché ci dovrebbero essere più probabilità di reperire una quantità maggiore di varianti per proverbi aventi un'alta frequenza d'uso. Ricordiamo, infatti, che per varianti di proverbi non intendiamo solamente le varianti che potremmo trovare nei dizionari, ma anche quei casi in cui i proverbi sono modificati, sono enunciati parzialmente oppure sono parafrasati, o ancora presentano l'aggiunta di parole, espressioni o gli stessi marcatori di proverbi al loro interno. Pertanto, aumentando i casi in cui un proverbio è utilizzato, aumenteranno, logicamente, anche le diverse maniere di enunciare lo stesso proverbio. In base anche ai nostri dati, si nota che non di rado ogni parlante fa un uso personale dei proverbi, modificandoli, storpiandoli o scegliendo di enunciarli in maniera non convenzionale o comunque in base a un proprio stile personale.

In secondo luogo, ricordiamo che uno studioso di grande statura e competenza come Anscombe (2005) si scaglia fortemente contro la tesi della fissità dei proverbi. Una delle principali ragioni di questa sua controtesi, ampiamente corroborata dal presente studio, riguarda la presenza di numerose varianti di proverbi e di proverbi modificati: se i proverbi possono essere così elastici, ciò significa che essi non sono fissi quanto, ad esempio, le espressioni idiomatiche. Pertanto, essendo così evidente e pervasiva la questione della *non fissità* dei proverbi, questo potrebbe significare due cose: da un lato, che la variabilità dei proverbi, ovvero la loro capacità di generare varianti, non dovrebbe avere alcun peso ed essere ininfluyente, e, dall'altro lato, che, siccome i proverbi non sono fissi, a una maggiore frequenza d'uso dovrà corrispondere una maggiore capacità di generare varianti.

Vediamo, ora, come abbiamo realizzato questa analisi. Prima di tutto, si è ribadita la validità del concetto di *indice di variabilità*, il solo a essere in grado di prendere in considerazione le varianti dei proverbi e le loro occorrenze totali. Se avessimo considerato solamente i valori assoluti delle varianti, avremmo ottenuto i dati pronosticati, in quanto è matematico, o quasi, che a un aumento della frequenza d'uso corrisponda un incremento del numero delle varianti; per noi, invece, sarebbe stato molto più interessante capire il rapporto tra occorrenze totali e numero di modifiche. Questa è l'unica maniera in cui è possibile confrontare proverbi aventi una frequenza totale di qualche decina di unità con proverbi che, invece, superano le 200 occorrenze, come succede nel nostro studio.

In secondo luogo, si è deciso di lavorare non tanto sui valori puntuali dei singoli proverbi, quanto piuttosto sulle medie di valori individuate per scaglioni di frequenze. Per suddividere i proverbi in base alla loro frequenza, abbiamo anzitutto separato i proverbi che mostravano una frequenza reale inferiore alle 20 occorrenze (che ricordiamo essere la soglia che abbiamo individuato per poter fornire una descrizione esauriente del proverbio) dal resto dei risultati, creando una categoria 0-19, ovvero composta da proverbi aventi una frequenza d'uso minore o uguale a 19. Successivamente, abbiamo considerato la quantità di proverbi per ogni decina e ogni centinaia di occorrenze, confrontando la distribuzione tra i proverbi di itWaC e quelli di frWaC per individuare gruppi bilanciati di occorrenze. In questa maniera siamo giunti alla seguente suddivisione dei proverbi. Il primo gruppo di questi è formato dai proverbi aventi da 20 a 39 occorrenze (che sono 19 per i proverbi di itWaC contro 7 per i proverbi di frWaC), seguito dal gruppo di proverbi da 40 a 59 occorrenze (composto da 8 proverbi italiani contro 7 proverbi francesi) e da quello da 60 a 99 occorrenze (contenente 16 proverbi italiani contro 9 proverbi francesi). Gli scaglioni successivi sono stati notevolmente allargati in quanto il numero dei proverbi per ognuno di questi andava via via diminuendo. Si è così individuato lo scaglione da 100 a 199 proverbi (di 12 proverbi, quello italiano, e 10 proverbi, quello francese) e, l'ultimo, oltre i 200 proverbi (composto da 13 proverbi italiani contro 6 proverbi francesi). Si noterà che, nonostante gli sforzi compiuti per cercare di giungere a categorie numericamente omogenee tra il corpus italiano e quello francese, rimangono comunque notevoli differenze numeriche. Ecco perché ci siamo avvalsi del calcolo della media per ogni scaglione di frequenza.

Presentiamo, dunque, i risultati nel prossimo grafico (Fig. 33). In esso, si illustrano simultaneamente i valori dei proverbi di itWaC e di quelli di frWaC.

Variabilità dei proverbi da itWaC e frWaC in funzione della frequenza effettiva

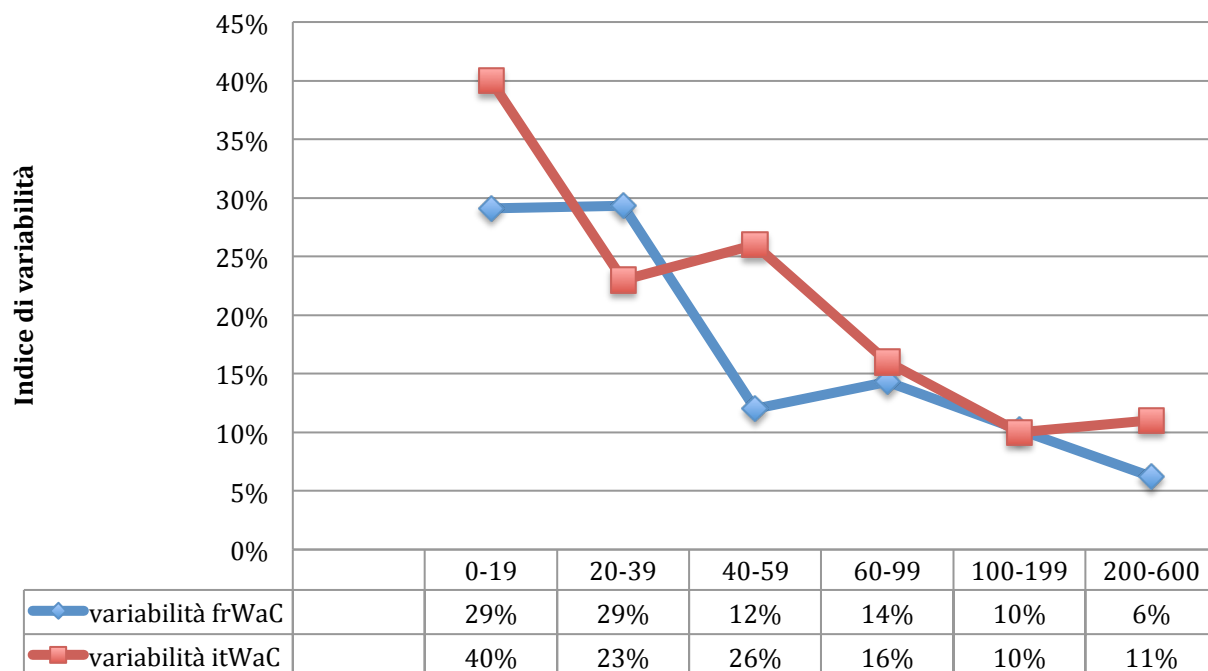


Fig. 33. Indici di variabilità dei proverbi di itWaC e frWaC in funzione della frequenza effettiva dei proverbi nei due corpora. In basso, nella tabella, sono riportati i valori medi degli indici di variabilità per ogni scaglione di frequenza indicato nella prima riga in alto della tabella. Tale riga funge anche da asse orizzontale del grafico.

Emerge una chiara tendenza: quella dell'abbassamento dell'indice di variabilità dei proverbi in funzione dell'innalzamento della loro frequenza. Si tratta, chiaramente, solo di una tendenza: vediamo, infatti, che ancora una volta non abbiamo una retta bensì una curva con andamento ondulatorio. Pensiamo ai valori di itWaC dei tre scaglioni compresi nell'intervallo da 0 a 49: si va da un indice di variabilità del 40% per il primo gruppo a un valore del 26% per il terzo gruppo, passando da una percentuale intermedia del 23%, pertanto inferiore ai due estremi, per il secondo scaglione. Abbiamo così un abbassamento per poi avere un piccolo aumento. Ciò significa che, in media, gli indici di variabilità dello scaglione da 40 a 59 occorrenze sono maggiori rispetto a quelli relativi ai proverbi aventi da 20 a 39 occorrenze. Anche per i dati estrapolati da frWaC si verifica questo movimento ondulatorio: lo scaglione da 60 a 99 è maggiore rispetto a quello da 40 a 59 e da 100 a 199. Tuttavia, è indubbio che all'aumentare delle occorrenze effettive dei proverbi diminuiscano gli indici di

variabilità. Ciò significa che più un proverbio è frequente e più è fisso o, quanto meno, genera una quantità inferiore di modifiche.

Cerchiamo di spiegare questo comportamento. Si potrebbe pensare che ciò avvenga in quanto le varianti o le modifiche sono possibili solo fino a un tetto massimo. Se così fosse, sarebbe logico che a una frequenza elevata corrisponda, in proporzione, una quantità ristretta di varianti. Ci siamo, quindi, concentrati sulle quantità medie di varianti per ogni scaglione determinato, senza metterle in rapporto con i valori delle occorrenze. Se si osserva il prossimo grafico (Fig. 34), si può notare che la quantità media di varianti reperite per scaglione di frequenza è direttamente proporzionale alle occorrenze dei proverbi.

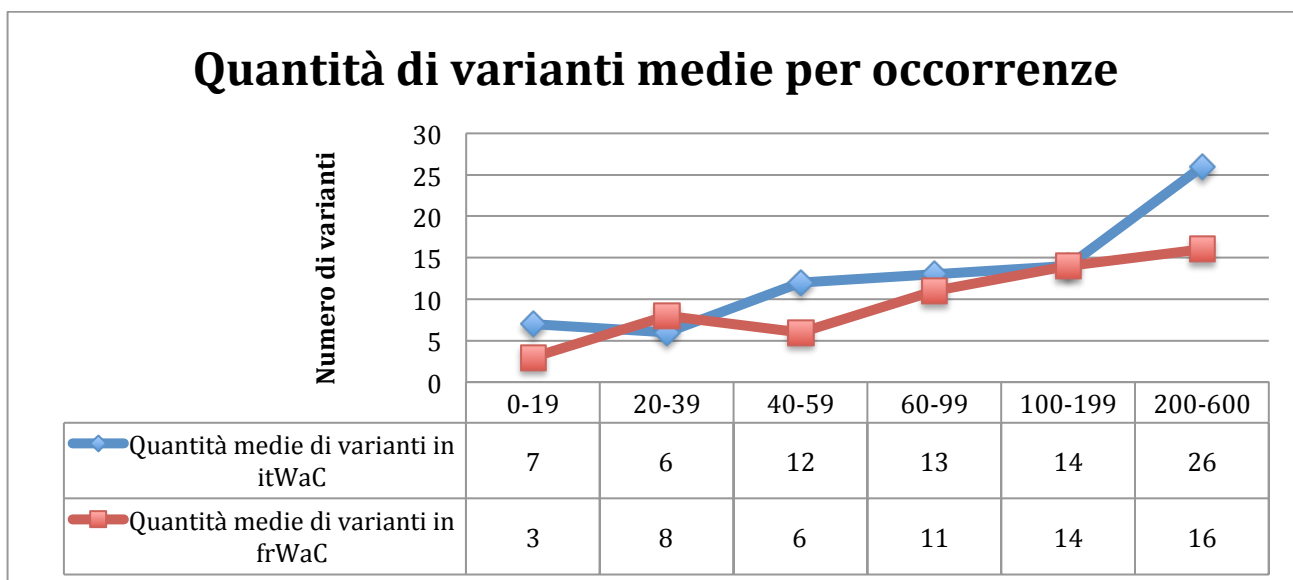


Fig. 34. Medie di varianti reali riscontrate in itWaC e frWaC.

Ciò significa che potenzialmente non c'è un tetto massimo al numero di varianti possibili. Ecco perché l'indice di variabilità che si è introdotto nel presente studio ha un grande valore: esso, tenendo conto del rapporto tra frequenza e numero di varianti, può farci comparare proverbi aventi frequenze molto diverse. Inoltre, il fatto che l'indice di variabilità diminuisca all'aumentare delle occorrenze impone di rivedere il concetto di fissità relativo ai proverbi.

Probabilmente, infatti, la fissità nei proverbi è più presente di quanto non si creda. Come già affermato (cfr. § 1.3 e 2.9), non si può assimilare la fissità del proverbio a quella tipica delle espressioni idiomatiche. Tale operazione non può essere compiuta in quanto permane una differenza tutt'altro che trascurabile tra proverbi ed espressioni idiomatiche: i

primi sono frasi mentre le espressioni idiomatiche sono costituenti. Ciò impone necessariamente diversi comportamenti sintattici e quindi tipologie altrettanto differenti di fissità.

Ciononostante, i dati mostrano che esiste un qualche tipo di fissità anche per i proverbi, altrimenti non si spiegherebbe perché più un proverbio è diffuso nella comunità linguistica e più è fisso.

A supporto di questa tesi, è possibile avanzare un'altra argomentazione. La possibilità di ottenere, con un'unica chiave di ricerca, una gamma di proverbi, sia in forma standard sia modificati, e di escludere automaticamente altri risultati che non siano proverbi, indica che esiste, anche per i proverbi, una qualche forma di fissità. Forse è questo il nodo da sciogliere: appurato che la fissità dei proverbi abbia caratteristiche assolutamente peculiari, bisogna studiare più approfonditamente questo ambito. Se, ad esempio, la ricerca *".*ra" [tag="Nc"] "e" [tag="Nc"] "non" [{}0,1} "mettere"¹¹²* ci fa giungere a un proverbio e alle sue varianti, ovvero, *Tra moglie e marito non mettere il dito, Tra medico e paziente non mettere un giudice, Fra moglie e marito non ci mettere Internet* oppure *Tra moglie e marito non ci mettere i suoceri*, ciò significa che la struttura utilizzata come chiave di ricerca è la struttura minima necessaria per fare giungere al maggior numero di proverbi di quel tipo. Si tratterà, pertanto, della struttura portante del proverbio, che, proprio come un muro portante, non può essere modificata per non compromettere la stabilità del proverbio. Ovviamente, in questa metafora, il muro portante, se viene modificato, non comprometterà il proverbio, ma impedirà di arrivare al proverbio stesso. In altre parole, non si arriverà più a quel proverbio, bensì a una gamma di frasi che potrebbe comprendere il proverbio, ma anche frasi libere che non hanno alcun rapporto con il proverbio stesso.

4.7. Ulteriori piste di ricerca

In questo capitolo, si sono delineate alcune delle piste di ricerca che abbiamo seguito non tanto per mostrare la validità di certe tesi o confutarne altre, quanto piuttosto per illustrare le potenzialità di una ricerca di questo tipo in ambito paremiologico, un campo in cui molto è ancora da scoprire, come dimostrano anche i dati in nostro possesso.

¹¹² Traduciamo in un linguaggio più comprensibile tale chiave di ricerca. *".*ra"* significa qualsiasi parola che finisce con le lettere *ra*. *[tag="Nc"]* significa qualsiasi parola che sia categorizzata come nome comune. *"e"* indica qualsiasi parola formata solamente dalla lettera *e* così come *"non"* indica una parola composta dalle lettere *n, o, n*. *[{}0,1}* significa che tra la parola precedente e quella successiva ci deve essere da zero a una parola di qualsiasi tipo. *"mettere"* indica, appunto, il verbo *mettere* alla forma infinita.

Qualsiasi ricerca linguistica sul proverbio deve però partire da un presupposto che oggi si fatica ad accettare: i proverbi non stanno scomparendo, ma sono tutt'ora utilizzati e creati; pertanto, la ricerca non può fermarsi. Cercheremo, ora, di dimostrare queste affermazioni.

In primis, se i proverbi fossero davvero in via di estinzione, non si capirebbe come mai esistano casi di creazione di nuovi proverbi. Vi sono esempi, come quelli da (62) a (64), in cui i parlanti indicano con l'etichetta proverbio una frase che effettivamente risponde a tutte le condizioni necessarie tipiche del proverbio, salvo a quella della notorietà.

- (62) Non è mia intenzione dirvi tutto su come l'open Source funziona: sarebbe come cercare di spiegare come funziona la lingua inglese. Qualcosa, però, posso dirvi del Perl e di dove sta andando. Un altro proverbio: Le tre grandi virtù del programmatore sono la pigrizia, l'impazienza e la tracotanza. I grandi programmatori in Perl le abbracciano tutte e tre, e così gli sviluppatori Open Source.¹¹³
- (63) Su Internet, dice il proverbio, nessuno sa che sei un cane. Puoi nasconderti dietro uno pseudonimo o il nome di un'altra persona, puoi cambiare sesso e identità, puoi smerciare falsi e patacche in barba ai diritti d'autore.¹¹⁴
- (64) (...) anche perché, alla peggio, come dice il proverbio del giornalista: "Una buona didascalia salva la faccia a qualsiasi fotografia!"¹¹⁵

In questi tre esempi, si trovano proverbi nuovi o comunque prodotti nella nostra società. Nell'esempio (62) si parla delle virtù che dovrebbe avere un programmatore di software: dato il mestiere, non può che trattarsi di un proverbio creato recentemente. Lo stesso dicasi per l'esempio (63), in cui il tema centrale è Internet, e per l'esempio (64), nel quale si parla di un altro mestiere, quello del giornalista, sicuramente meno recente di quello del programmatore, ma che è comunque insolito trovare nei proverbi tradizionali. Le stesse considerazioni valgono anche per alcuni esempi, tratti da frWaC, che sono proposti da (65) a (69): si tratta di proverbi relativi ad argomenti molto affini alla nostra società e che non si trovano facilmente nei proverbi tradizionali.

¹¹³ URL: <<http://www.flavioweb.it/HackerStory/OpenSources/larrywall.html>>.

¹¹⁴ URL: <<http://lgxserver.uniba.it/lei/rassegna/981029.htm>>.

¹¹⁵ URL: <http://digilander.libero.it/ubuland/frames/artic_freak.html>.

- (65) Il nous faut donc encore patienter et continuer de guetter les montagnes. Car comme le dit un autre vieux proverbe : "Un skieur de rando est toujours à l'affût des bonnes conditions".¹¹⁶
- (66) Fin mai, début juin, période d'examen, dit le bien connu proverbe...¹¹⁷
- (67) Quand Wall Street éternue, Paris s'enrhume (proverbe boursier).¹¹⁸
- (68) "Le bon programmeur n'est pas celui qui ne fait pas d'erreur, c'est celui qui les prévoit" (proverbe silicon-valleyen).¹¹⁹
- (69) Un vieux proverbe vigneron dit « il y a trompe - l'oeil mais il n'y a pas trompe goût »... Pierre Laforest permet de déguster et apprécier nos vins dans votre environnement familial et bien sûr, si vous le souhaitez, en compagnie de vos amis et proches (...).¹²⁰

Il fatto che si tratti di proverbi nuovi è confermato dall'impossibilità di trovare gli stessi proverbi nei dizionari di proverbi da noi consultati¹²¹ o dalle scarse occorrenze degli stessi cercati in Internet.

Ora, non stiamo dicendo che bisogna considerare queste frasi veri e propri proverbi: sebbene gli autori le mostrino come tali, non possiamo non tenere in considerazione la variabile essenziale della notorietà della frase affinché possa essere considerata un proverbio. Ciononostante, siamo anche consapevoli del fatto che il processo necessario perché una frase si trasformi in proverbio richieda molto tempo (sebbene nessuno abbia mai quantificato con certezza e precisione tale lasso temporale). Pertanto, possiamo sostenere che le frasi presentate negli esempi da (62) a (69) siano candidate a divenire proverbi: solo il tempo e l'uso potranno confermare o smentire tale pronostico. Il fatto, però, che ci sia ancora interesse a definire nuove frasi con l'etichetta di proverbio, mostra che la categoria dei proverbi, appunto, è ben lungi dall'essere esaurita o vicina alla sua fine.

Un'altra prova a sostegno della vitalità dei proverbi risiede nella quantificazione dei proverbi effettivamente in uso. Ora, non è facile fornire dati certi relativi alla loro frequenza d'uso. Siamo consapevoli del fatto che il nostro lavoro sia stato alquanto selettivo nei confronti dei proverbi da analizzare, il che offre un panorama ristretto di questi elementi.

¹¹⁶ URL: <<http://www.skitour.fr/blog/neigedecorse/688-corse-la-sous-couche-est-arrivee>>.

¹¹⁷ URL: <http://www.gchagnon.fr/daggerfall/lettre/n40_05_2004.html>.

¹¹⁸ URL: <http://pointsdactu.bm-lyon.fr/article.php3?id_article=565>.

¹¹⁹ URL: <<http://www710.univ-lyon1.fr/%7efouet/DEUG1/chap13.html>>.

¹²⁰ URL: <<http://www.pierre-laforest.fr/index.php?ID=42>>.

¹²¹ Cfr. § 3.7 per la lista e le caratteristiche principali dei dizionari da noi utilizzati.

Ciononostante, abbiamo voluto cercare, almeno per la parte relativa a itWaC, tutti i proverbi ottenuti nella prima fase dello studio all'interno dell'intero corpus. L'obiettivo era di ottenere la media matematica delle occorrenze totali per ogni proverbio ottenuto nella fase iniziale. Abbiamo così cercato la totalità dei 617 proverbi italiani: per quelli già cercati nella seconda fase dello studio, abbiamo riutilizzato i dati in nostro possesso, mentre per quelli non ancora cercati, di occorrenza iniziale da 1 a 3, abbiamo impostato come chiave di ricerca la variante più frequente oppure l'unica variante presente, o ancora, in caso di parità, quella che fa giungere al numero maggiore di risultati. Abbiamo così individuato il numero medio di occorrenze per ogni proverbio: tale valore è pari a 69. Ciò significa che ogni proverbio può avere una media di 69 occorrenze nel corpus. Moltiplicando tale valore per la quantità totale dei nostri proverbi italiani (che, ricordiamo, escludono gran parte dei risultati ottenuti, specialmente quelli relativi ai “proverbi non italiani” oppure relativi a quei casi in cui mancava un vero contesto enunciativo, etichettati con il nome convenzionale di “interesse storico-folkloristico”) abbiamo una frequenza d'uso minima di proverbi all'interno di itWaC pari a 42.573. Per avere un'idea della percentuale d'uso dei proverbi sulla totalità del corpus, si è deciso di prendere in considerazione non tanto la quantità totale delle parole del corpus, quanto piuttosto la quantità totale delle frasi presenti in itWaC. Questo perché, avendo constatato che il proverbio è necessariamente una frase, ha certamente più senso confrontarlo con una grandezza comparabile. Il totale delle frasi in itWaC è di 68.147.599, incluse le ripetizioni che, invece, abbiamo completamente escluso dai nostri calcoli sulle occorrenze dei proverbi. Una maniera per ovviare al problema delle ripetizioni all'interno del corpus è quello di cercare di toglierle dal numero totale di frasi presenti. Ci possiamo rifare ai dati che abbiamo individuato in § 3.2.6 relativo alle ripetizioni riscontrate tra i nostri risultati: tale dato era pari al 18%. Potremmo pertanto togliere dai 62 milioni di frasi il 18% circa. Il totale sarebbe, approssimativamente, di 56 milioni. I nostri proverbi ricoprirebbero non meno dello 0,08% delle frasi presenti in itWaC (considerando che i nostri dati si riferiscono alla quantità minima di proverbi riscontrabile nel corpus). Un dato apparentemente non molto elevato, ma se comparato con qualsiasi altra frase ci fa capire quanto sia tutt'altro che trascurabile.

4.7.1. Proverbi e dizionari: come migliorare questo binomio

Appurato che i proverbi siano tutt'ora utilizzati e creati, passiamo ad alcune delle principali piste di ricerca che si potrebbero percorrere in futuro.

Per prima cosa, si potrebbero rivedere i classici dizionari di proverbi e magari ripensare alla loro funzione o ai principi di costruzione. Si è già dimostrato in § 4.5 che in diverse occasioni i dizionari si dimostrano non all'altezza del loro compito: il fatto che non prendano in considerazione certi proverbi aventi alte frequenze d'uso è una grave mancanza.

Ciononostante, bisognerebbe anche migliorare la loro concezione, specialmente per quanto attiene ai metodi di ricerca dei proverbi: si è più volte indicata l'inadeguatezza della parola chiave per reperire i proverbi nei dizionari (cfr. § 3.7) o degli indici tematici. Riguardo a quest'ultimo sistema di categorizzazione dei proverbi, dobbiamo rilevare la troppa soggettività rivestita dalle categorie tematiche e, al contempo, la loro inadeguatezza. Per capire ciò, concentriamoci sulle definizioni dei proverbi fornite dai dizionari, dalle quali scaturisce, in diverse occasioni, la categorizzazione tematica dei proverbi.

Prendiamo, ad esempio, il proverbio *Moglie e buoi dei paesi tuoi*. Il dizionario di Guazzotti e Oddera (2006: 316) dà la seguente definizione: "è bene che il matrimonio e gli affari avvengano fra persone dello stesso ambiente". Tuttavia, nel nostro corpus abbiamo reperito alcuni esempi di questo proverbio modificato: *Donne e buoi dei paesi tuoi*; *Manager e buoi dei paesi tuoi*; *Gatti e buoi dei paesi tuoi*; *Inter e buoi dei paesi tuoi*; *Mucche e buoi dei paesi tuoi*. In questi esempi, notiamo che i riferimenti al matrimonio oppure agli affari vengono meno, mentre affiorano altri significati probabilmente più vicini al significato paremiologico del proverbio. Il grande denominatore concettuale comune di tutti questi esempi riguarda, banalmente, il fatto che sia necessario circondarsi di persone, collaboratori, amici, anche in senso lato (comprendendo gli animali), o comunque che sia meglio preferire ciò che ha la nostra stessa origine. Il *matrimonio* e gli *affari* non riguardano il significato paremiologico del proverbio, bensì quello compositivo. In altri termini, i dizionari dovrebbero preoccuparsi maggiormente di estrapolare il significato paremiologico nella fase di categorizzazione e definizione dei proverbi.

4.7.2. Variabilità e semantica dei proverbi

Anche la questione della variabilità dei proverbi merita ulteriori ricerche e riflessioni. Si tratta di un tema di grande interesse per il funzionamento semantico dei proverbi e, probabilmente, di uno dei modi per riuscire a decifrare i significati paremiologici spesso confusi con i significati compositivi.

In primo luogo, è necessario approfondire e migliorare le tecniche di estrazione di tutte le varianti dei proverbi, comprese le versioni modificate degli stessi.

In secondo luogo, si devono operare serie riflessioni sulle strutture che più di tutte rimangono fisse, non solo dal punto di vista semantico, ma anche morfosintattico. Da qui scaturisce l'importanza di uno studio *corpus-driven* relativo alla questione della fissità dei proverbi, che deve essere ristudiata e ripensata senza preconcetti ma alla luce delle evidenze derivate dai corpora. Accanto a tale questione, bisognerebbe analizzare anche il tema della *proverbialisation* e della *déproverbialisation*: quali sono le tappe perché una frase si possa affermare come proverbio e che cosa succede, nella pratica e a tutti i livelli, quando una frase viene modificata?

Per fare ciò, nonché per approfondire l'intera ricerca e le altre questioni ancora aperte, sono necessari ulteriori dati e ulteriori tecnologie. Si noterà, ad esempio, che non sono stati presi in considerazione i dati sul parlato, in quanto non contemplati dai corpora WaCky. Tali dati, invece, potrebbero fornire ulteriori dettagli per risolvere molte delle questioni ancora aperte.

La ricerca non si può fermare, specialmente in un ambito in continua evoluzione come quello dei proverbi: ormai le tecnologie e i dati esistono, serve solo la costanza e l'impegno della comunità scientifica, nonché la volontà di ammettere che i proverbi siano troppo importanti, dal punto di vista linguistico, culturale e comunicativo (pensiamo a quanto vengano sfruttati nella pubblicità, ambito di cui, purtroppo, non ci siamo quasi per nulla interessati nella presente tesi), per essere giudicati inutili oggetti polverosi.

Conclusioni

Non è semplice trovare una conclusione per un lavoro così ampio e complesso come quello presentato nella Tesi di Dottorato di cui ci siamo occupati. La ricerca in oggetto ha aperto, infatti, più interrogativi di quanti non ne abbia risolti. La sensazione, per chi ha vissuto in prima persona tale progetto, è che più si avanzava nello studio e più emergevano nuove questioni da risolvere.

La ricerca è partita da una prima fase di analisi e critica degli studi esistenti in ambito paremiologico. Nel primo capitolo, sono stati mostrati i nodi cruciali dello studio linguistico sul proverbio. *In primis*, si è cercato di fornire una definizione di proverbio, sottolineando che, ancora oggi, non vi è un'unica definizione ampiamente accettata, bensì tante definizioni che differiscono per diversi aspetti. Siamo partiti da una considerazione di estrema importanza: il proverbio è una frase che può occasionalmente trasformarsi in un insieme di frasi, vale a dire in un testo, ma che vede nella frase la soglia minima al di sotto della quale non può scendere senza perdere il proprio status di proverbio. Da questa osservazione scaturisce l'impossibilità di confrontare i proverbi con le espressioni idiomatiche: se i primi sono frasi, le espressioni idiomatiche sono costituenti di frasi.

In secondo luogo, si è tentato di approfondire l'analisi linguistica del proverbio proponendo un nuovo paradigma definitorio: i proverbi sono frasi *ON-sentencieuses* (cfr. § 1.4.1), nonché frasi generiche (cfr. § 1.4.2) in grado di denominare stati di cose generali, abituali e consuetudinari (cfr. § 1.4.3).

In seguito, è stata analizzata la semantica del proverbio suddividendola in tre piani distinti: il significato compositivo, ossia determinato dalle parole in esso presenti, il significato paremiologico e il messaggio veicolato. In particolare, ci siamo soffermati sul significato paremiologico: a differenza di qualunque frase generica, il proverbio può essere interpretato in senso umano, il che non sarebbe possibile per una qualunque frase generica come *L'oro è giallo* (cfr. § 1.4.3).

In § 1.4.4, ci siamo interrogati sul rapporto tra i tre piani del significato del proverbio. Il proverbio sembra mostrare un andamento parabolico: parte da un significato compositivo che fa giungere a un significato paremiologico, attraverso un rapporto di carattere metaforico e metonimico, che ha un grado di astrazione molto più elevato rispetto al significato compositivo. Da questo livello di astrazione, il significato torna ad assumere un carattere di specificità quando entra nella comunicazione e si trasforma in messaggio: esso,

infatti, denota una situazione particolare direttamente determinabile dal contesto della situazione comunicativa in cui viene enunciato, mantenendo intatta la relazione con la situazione generica di cui è denominazione a livello di significato paremiologico, ma applicando tale struttura alle situazioni specifiche che emergono dai contesti contingenti.

In § 1.5 sono stati illustrati fenomeni estremamente importanti e frequenti per l'analisi del proverbio come la *déproverbialisation*, ossia la perdita puntuale e occasionale dello status di proverbio da parte di una frase, e il *détournement*, ovvero le modifiche apportate ai proverbi. Infine, in § 1.6, ci si è concentrati sul funzionamento dei proverbi nel discorso, da cui sono emersi, tra gli altri, fattori di grande importanza per il nostro studio, come le espressioni che introducono solitamente i proverbi.

Nel secondo capitolo, si è passati dall'ambito linguistico legato alla paremiologia all'ambito della *corpus linguistics* applicata alla fraseologia e, specialmente, alla paremiologia.

Dopo aver spiegato perché si è scelto di studiare i proverbi applicando i principi della linguistica dei corpora (cfr. in particolare § 2.1 e 2.6), abbiamo presentato i due corpora utilizzati in questa ricerca, vale a dire itWaC e frWaC, analizzandone le caratteristiche principali e dimostrandone la piena compatibilità e comparabilità. Successivamente, abbiamo analizzato l'approccio *corpus-driven* (cfr. § 2.7) scelto per effettuare questo studio: tale procedimento permette di giungere a determinate ipotesi ignote a priori esclusivamente attraverso l'analisi di fatti ricorrenti. Nell'ultima parte del capitolo, abbiamo illustrato le tappe necessarie per formulare un metodo di ricerca *corpus-driven* dei proverbi, al fine di ottenere automaticamente un campione di proverbi. Per fare ciò, abbiamo esaminato, in § 2.8, alcune esperienze di applicazione della linguistica dei corpora alla fraseologia.

Nel terzo capitolo, sono state dettagliatamente illustrate le tappe di costituzione del *database* di proverbi italiani e francesi tramite l'analisi dei corpora itWaC e frWaC. Sono stati, dapprima, presentati i risultati generali, ossia le concordanze dei marcatori di proverbio, per poi passare alle fasi di scrematura di tali concordanze, al fine di costituire la vera e propria banca dati dei nostri proverbi. Il frutto principale di questa fase sono stati i 617 proverbi italiani e 630 proverbi francesi individuati in maniera semiautomatica ed effettivamente attestati nei due corpora.

Il quarto capitolo si è focalizzato sull'analisi dei dati: dopo aver determinato una lista di frequenza dei proverbi, sono stati individuati i criteri per analizzare i proverbi stessi all'interno dei due corpora senza utilizzare alcun marcatore di proverbio. Questo capitolo ha

fornito l'occasione per fare il punto della situazione sullo studio dei proverbi. In base ai dati in nostro possesso, è emersa la necessità di studiare le varianti dei proverbi, il che ha portato alla formulazione di un *indice di variabilità* in grado di esprimere la capacità di un proverbio a generare proverbi modificati in rapporto alle occorrenze totali dello stesso proverbio. Tale indice ha mostrato quanto, tendenzialmente, siano i proverbi meno modificati, e quindi più fissi, ad avere le maggiori frequenze d'uso, il che impone di riconsiderare il tema della non fissità del proverbio. Sono state altresì evidenziate questioni di grande importanza, come la competenza paremiologica dei parlanti o il rapporto tra dizionari e proverbi o, ancora, il fatto che i proverbi siano più utilizzati di quanto non si creda comunemente. Infine, si è dedicata una breve sezione (§ 4.7) ai punti principali che devono essere necessariamente sviluppati in future ricerche.

La sensazione, alla luce della ricerca in oggetto, è che il proverbio sia stato per troppo tempo relegato a un ruolo di polveroso oggetto folklorico, appartenente a un passato ormai tramontato, e che non siano mai stati presi adeguatamente in considerazione i suoi aspetti linguistici più problematici. Come abbiamo più volte ribadito nel corso della presente Tesi di Dottorato, i proverbi continuano a essere creati, utilizzati e prolifici più di quanto non si creda. Per questa ragione la ricerca necessita di ulteriori approfondimenti, avvalendosi anche delle nuove tecnologie.

Probabilmente, la conclusione di questo lavoro è rappresentata dalla consapevolezza che il proverbio sia ancora troppo poco conosciuto da non essere più studiato. La linguistica dei corpora, come dimostrato in questa ricerca, può permettere di raggiungere nuovi traguardi in quanto è in grado non solo di mostrare dati a oggi ancora ignoti, ma anche di trattarli e analizzarli in maniere inedite. Questo funge da ponte con un obiettivo ancora molto lontano: fornire nuovi strumenti, in ambito paremiologico ma non solo, di categorizzazione linguistica. Poter giungere a un dizionario di proverbi in cui essi vengano categorizzati esclusivamente in base al loro significato paremiologico sarebbe, ad esempio, una conquista non solo per la paremiologia ma anche, in generale, per la linguistica. Poter astrarre il significato paremiologico sulla base di un insieme di proverbi aventi un'origine in comune (un proverbio) ma diverse declinazioni (le varianti del proverbio) potrebbe aiutare a realizzare tale obiettivo, non solo in senso intralinguistico, ma anche in maniera interlinguistica. Sarebbe un'enorme conquista poter fornire un dizionario capace di mostrare tutti i proverbi, in una o

più lingue, aventi lo stesso significato paremiologico ma diverse declinazioni sul piano del significato compositivo.

La ricerca, come emerge chiaramente, è lungi dall'essere terminata. L'obiettivo principale di questo studio è stato quello di aver gettato le basi affinché essa possa continuare a produrre risultati sempre più innovativi, in quanto sempre più legati alla realtà linguistica e alle esigenze di coloro che ne beneficeranno, che non saranno solamente i linguisti o i paremiologi, ma anche i professionisti delle lingue o chiunque sia coinvolto nell'apprendimento o nell'approfondimento di una lingua.

Bibliografia

- Agostiniani, L. (1978). Semantica e referenza nel proverbio. In: *Archivio Glottologico Italiano*, vol. LXIII, pp. 78-109.
- Anscombe, J.-C. (1989). Théorie de l'argumentation, topoï et structures discursives. In: *Revue Québécoise de Linguistique*, vol. 18, n. 1, pp. 13-56.
- Anscombe, J.-C. (1997). Reflexiones críticas sobre la naturaleza y el funcionamiento de las paremias. In: *Paremia*, n. 6, pp. 43-54.
- Anscombe, J.-C. (2000). Parole proverbiale et structures métriques. In: *Langages*, vol. 34, n. 139, pp. 6-26.
- Anscombe, J.-C. (2005). Les proverbes : un figement du deuxième type ? In: *Linx* [online], vol. 53, pp. 17-33. URL: <<http://linx.revues.org/255>>. (Consultato l'11 ottobre 2012).
- Baker, P. (a cura di) (2009). *Contemporary Corpus Linguistics*. London, New York: Continuum.
- Balibar-Mrabti, A. (2005). Semi-figement et limites de la phrase figée. In: *Linx* [online], vol. 53, pp. 35-54. URL: <<http://linx.revues.org/255>>. (Consultato l'11 ottobre 2012).
- Barbadillo de la Fuente, M. T. et al. *Refranero Multilingüe*. [online]. <http://cvc.cervantes.es/lengua/refranero/Default.aspx> (consulté le 10.01.2013).
- Baroni, M., Bernardini, S., Ferraresi, A. ed E. Zanchetta (2009). The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora. In: *Language Resources and Evaluation* [on-line], vol. 43, n. 3, pp. 209-226. URL: <<http://wacky.sslmit.unibo.it/doku.php?id=publications>>. (Consultato il 9 ottobre 2014).
- Barsanti Vigo, M. J. (2006). Problemática en torno al refrán y otras categorías paremiales. In: *Diccionarios y fraseología. Anexos de Revista de Lexicografía*, n. 3, pp. 197-206.
- Biber, D. (1993). Representativeness in Corpus Design. In: *Literary and Linguistic Computing*, vol. 8, n. 4, pp. 243-257. Oxford: Oxford University Press.
- Bidaud, F. (2002). *Structures figées de la conversation. Analyse contrastive français-italien*. Bern et al.: Lang.

- Boggione, V. (2004). Lógos, dialogo, letteratura. In: Boggione, V. e Massobrio, L. *Dizionario dei proverbi*, pp. XXIII-XLVIII. Torino: Utet.
- Boggione, V. e L. Massobrio (2004). *Dizionario dei proverbi*. Torino, Utet.
- Bronzini, G. B. (1999). La logica del proverbio: problemi e prospettive di classificazione e analisi. In: Trovato, S. C. (a cura di). *Proverbi, locuzioni, modi di dire nel dominio linguistico italiano. Atti del I Convegno di Studi dell'Atlante Paremiologico Italiano (API). Modica, 26-28 ottobre 1995*. Roma: Il Calamo, pp. 45-56.
- Cabré, M. T. (2000). Sur la représentation mentale des concepts. In: *Le sens en terminologie*. Bejoint H. et Thoiron P. (a cura di). Lyon: Presses universitaires de Lyon. pp. 20-39.
- Campos, J. G. e A. Barella (1995). *Diccionario de refranes*. Madrid: Espasa Calpe.
- Canellada, M. J. e B. Pallares (2001). *Refranero español. Refranes, clasificación, significación y uso*. Madrid: Castalia.
- Cardey, S. et al. (2005). Langues et cultures, systèmes et traduction. In: *Meta* [online], vol. 50, n.4. URL: <<http://id.erudit.org/iderudit/019833ar>>. (Consultato il 29 aprile 2013).
- Casadei, F. (1996). *Metafore ed espressioni idiomatiche: uno studio semantico sull'italiano*. Roma: Bulzoni.
- Chiari, I. (2007). *Introduzione alla linguistica computazionale*. Roma: GLF Editori Laterza.
- Correas, G. (1992). *Vocabulario de refranes y frases proverbiales y otras fórmulas comunes de la lengua castellana en que van todos los impresos antes y otra gran copa que juntó el Maestro Gonzalo Correas*. Madrid: Visor.
- Cram, D. (1983). The linguistic status of the proverb. In: *Cahiers de lexicologie. Revue internationale de lexicologie et de lexicographie*. Vol. XLVIII (II), pp. 53-71.
- Cristilli, C. (1989). Il proverbio come esempio di testualità popolare. In: Vallini, C. *La pratica e la grammatica. Viaggio nella linguistica del proverbio*. Napoli: Istituto Universitario Orientale, Dipartimento di studi letterari e linguistici dell'Occidente, pp. 177-206.
- Deiana, R. (2001). *Invito a cena con aforisma. 300 citazioni e boutade d'autore sul cibo*. Torino: Il leone verde.

- Deignan, A. (2009). Searching for Metaphorical Patterns in Corpora. In: Baker, P. (a cura di). *Contemporary Corpus Linguistics*. London; New York: Continuum, pp. 9-31.
- Depecker, L. (2011). Comment aborder le concept d'un point de vue linguistique ? In: Briu, J.-J. *Terminologie (I) : analyser des termes et des concepts*. Bern: Lang, pp. 17-32.
- Dournon, J.-Y. (1986). *Dictionnaire des proverbes et dictons de France*. Paris: Hachette.
- Duneton, C. e S. Claval (1990). *Le bouquet des expressions imagées : encyclopédie thématique des locutions figurées de la langue française*. Paris: Seuil.
- Dylan, G. e K. Fischer (2010). *Quantitative Methods in Cognitive Semantics : Corpus-Driven Approaches*. Berlin; New York: De Gruyter Mouton.
- Greimas, A. J. (1986). *Sémantique structurale, recherche de méthode*. Paris: PUF.
- Gross, G. (1988). Degré de figement des noms composés. In: *Langages*, vol. 23, n. 90, pp. 57-72.
- Gross, G. (1996). *Les expressions figées en français*. Gap, Paris: Ophrys.
- Guazzotti, P. et Oddera, M. F. (2010). *Il grande dizionario dei proverbi italiani*. Bologna: Zanichelli.
- Halliday, M. A. K. (1992). Language Theory and Translation Practice. In: *Rivista internazionale di tecnica della traduzione*, n. 0, pp. 15-25.
- Honeck, R. P. (1997). *A proverb in mind: the cognitive science of proverbial wit and wisdom*. Mahwah (N. J.); London: Lawrence Erlbaum associates.
- Junceda, L. (1991). *Del dicho al hecho*. Barcelona: Obelisco.
- Kleiber, G. (1990). *La sémantique du prototype. Catégories et sens lexical*. Paris: PUF.
- Kleiber, G. (1999). Les proverbes : des dénominations d'un type « très très spécial ». In: *Langue française*, n. 123, pp. 52-69.
- Kleiber, G. (2000). Sur le sens des proverbes. In: *Langages*, vol. 34, n. 139, pp. 39-58.
- Lakoff, G. e M. Johnson (1980 [2003]). *Metaphors we live by*. Chicago; London: The University of Chicago Press.
- Lakoff, G. e M. Turner (1989). *More than cool reason: a field guide to poetic metaphor*. Chicago; London: The University of Chicago Press.

- Lapucci, C. (1993). *Dizionario dei modi di dire della lingua italiana*. Milano: Garzanti.
- Lapucci, C. (2006). *Dizionario dei proverbi italiani*. Firenze: Le Monnier.
- Leech, G. (1991). The State of the Art in Corpus Linguistics. In: Aijmer, K. e B. Aitenberg (a cura di). *English Corpus Linguistics. Studies in Honour of Jan Svartvik*. London, New York: Longman, pp. 8-29.
- Longhi, J. (2008). *Objet discursif et doxa. Essai de sémantique discursive*. Paris: L'Harmattan.
- Maloux, M. (1980 [2009]). *Dictionnaire des proverbes, sentences et maximes*. Paris: Larousse.
- Mejri, S. (2005). Figement absolu ou relative : la notion de degré de figement. In: *Linx* [online]. vol. 53, pp. 183-196. URL: <<http://linx.revues.org/283>>. (Consultato l'11 ottobre 2012).
- Michaux, C. (1996). Proverbes et jugement individuel : deux incompatibles. Projet ARC "Typologie textuelle et théorie de la signification". In: *Rapport de recherches*, n° 4. Bruxelles: Université Libre de Bruxelles.
- Michaux, C. (1999). Proverbes et structures stéréotypées. In: *Langue française*, n. 123, pp. 85-104.
- Montreynaud, F., Pierron, A. e F. Suzzoni (1989). *Dictionnaire de proverbes et dictons*. Paris: Dictionnaires Le Robert.
- Moon, R. (1998). *Fixed Expressions and Idioms in English. A Corpus-based Approach*. Oxford: Clarendon Press.
- Norricks, N. R. (1985). *How proverbs mean: semantic studies in English proverbs*. Berlin: Mouton.
- Paczolay, G. (2015). *Többnyelvű szólás- és közmondásgyűjtemény. 340 közmondás magyar, angol, észt, francia, német, olasz, lengyel, portugál, spanyol és latin nyelven*. Budapest: Tinta Könyvkiadó.
- Pamies Beltrán, A., Luque Durán, J. D. e Patricia Fernández Martín (a cura di) (2011). *Paremiología y herencia cultural*. Granada: Granada Lingvistica.

- Perrin, L. (2000). Remarques sur la dimension générique et sur la dimension dénomminative des proverbes. In: *Langages*, vol. 34, n. 139, pp. 69-80.
- Philip, G. (2011). *Colouring meaning: collocation and connotation in figurative language*. Amsterdam; Philadelphia: Benjamins.
- Pittano, G. (1996). *Frase fatta capo ha: dizionario dei modi di dire, proverbi e locuzioni*. Bologna: Zanichelli.
- Planelles, G. (2005). *Les expressions françaises décortiquées* [online]. URL: <<http://www.expressio.fr/>>. (Consultato il 10 gennaio 2013).
- Pocetti, P. (1989). Aspetti della teoria e della prassi del proverbio nel mondo classico. In: Vallini, C. *La pratica e la grammatica. Viaggio nella linguistica del proverbio*. Napoli: Istituto Universitario Orientale, Dipartimento di studi letterari e linguistici dell'Occidente, pp. 61-85.
- Polguère, A. (2008). *Lexicologie et sémantique lexicale : notions fondamentales*. Montréal: Les Presses de l'Université de Montréal.
- Prandi, M. (2004). *The Building Blocks of Meaning. Ideas for a philosophical grammar*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Prandi, M. (2006). *Le regole e le scelte. Introduzione alla grammatica italiana*. Torino: UTET.
- Quartu, B. M. (2001). *Dizionario dei modi di dire della lingua italiana: 10.000 modi di dire ed estensioni figurate in ordine alfabetico per lemmi portanti e campi di significato*. Milano: Rizzoli.
- Quiroga, P. (2006). *Fraseología italo-española. Aspectos de lingüística aplicada y contrastiva*. Granada: Granada lingüística.
- Rey, A. (1989). Préface. In: *Dictionnaire de proverbes et dictons*. Paris: Dictionnaires Le Robert, pp. IX-XVIII.
- Roche, C. (2007). Le terme et le concept : fondements d'une ontoterminologie. In: *TOTh 2007 (Terminologie & Ontologie : Théories et Applications)*. Annecy, 1er juin 2007, pp. 1-22.

- Roche, C. (2012). Ontoterminology: how to unify terminology and ontology into a single paradigm. LREC 2012. In: *Eighth International Conference on Language Resources and Evaluation*. Istanbul, 21-27 mai 2012 [online]. URL: <http://www.lrec-conf.org/proceedings/lrec2012/pdf/567_Paper.pdf>. (Consultato il 1 febbraio 2013).
- Rozan, J.-F. (1984). *La prise de notes en interprétation consécutive*. Genève: Librairie de l'Université Georg & Cie S. A.
- Schapira, C. (1999). *Les stéréotypes en français. Proverbes et autres formules*. Paris: Ophrys.
- Sevilla Muñoz, J. (2011). Aspectos culturales y transculturales de la paremiología: el mínimo paremiológico. In: Pamies Beltrán, A., Luque Durán, J. D. e P. Fernández Martín (a cura di) *Paremiología y herencia cultural*. Granada: Granada lingvistica, pp. 75-86.
- Shapira, C. (2000). Proverbe, proverbialisation et déproverbialisation. In: *Langages*, vol. 34, n. 139, pp. 81-97.
- Silvestri, D. (1989). Osservazioni in margine ai proverbi sumerici: strutture linguistiche e architetture testuali. In: Vallini, C. *La pratica e la grammatica. Viaggio nella linguistica del proverbio*. Napoli: Istituto Universitario Orientale, Dipartimento di studi letterari e linguistici dell'Occidente, pp. 11-29.
- Simpson, J. (1992). *The Concise Oxford Dictionary of Proverbs*. Oxford: Oxford University Press.
- Sinclair, J. (2003). *Reading concordances: an introduction*. Harlow: Longman.
- Sinclair, J. McH. (1996). The search for units of meaning. In: *Textus*, vol. 9, n. 1, pp. 71-106.
- Sinclair, J. McH. (2005). Corpus and text – Basic principles. In: Wynne, M. (a cura di). *Developing linguistic corpora: A guide to good practice*. Oxford: Oxbow Books, pp. 1-16.
- Stefanowitsch, A. e S. T. Gries (a cura di) (2006). *Corpus-based approaches to metaphor and metonymy*. Berlin; New York: Mouton de Gruyter.
- Taylor, J. R. (1994 [2003]). *La categorizzazione linguistica: i prototipi nella teoria del linguaggio*. Stefania Giannini (a cura di). Macerata: Quodlibet.
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
- Turrini, G. et al. (1999). *Capire l'antifona*. Bologna: Zanichelli.

- Vietri, S. (1985). *Lessico e sintassi delle espressioni idiomatiche. Una tipologia tassonomica dell'italiano*. Napoli: Liguori Editore.
- Visetti, Y. M. & Cadiot, P. (2006). *Motifs et proverbes. Essai de sémantique proverbiale*. Paris: Presses Universitaires de France.
- Xatara, C. M. et al. (2009). *Dictionnaire d'expressions idiomatiques: Français – Portugais – Français*. [online]. URL: <http://www.cnrtl.fr/dictionnaires/expressions_idiomatiques/>. (Consultato il 10 gennaio 2016).

Corpora utilizzati: itWaC e frWaC

frWaC (French Web). URL: <http://nl.ijs.si/noske/wacs.cgi/first?corpname=frwac&reload=1&iquery=&queryselector=cqlrow&lemma=&lpos=&phrase=&word=&wpos=&char=&cql=&default_attr=word&fc_lemword_window_type=both&fc_lemword_wsize=5&fc_lemword=&fc_lemword_type=all&fc_pos_window_type=both&fc_pos_wsize=5&fc_pos_type=all&usesubcorp=&fsca_text.url=&fsca_text.domain=&fsca_text.wordcount=>>.

itWaC (Italian Web). URL: <http://nl.ijs.si/noske/wacs.cgi/first?corpname=itwac&reload=1&iquery=&queryselector=cqlrow&lemma=&lpos=&phrase=&word=&wpos=&char=&cql=&default_attr=word&fc_lemword_window_type=both&fc_lemword_wsize=5&fc_lemword=&fc_lemword_type=all&fc_pos_window_type=both&fc_pos_wsize=5&fc_pos_type=all&fsca_text.url=&fsca_text.domain=&fsca_text.wordcount=>>.

APPENDICE

ALLEGATO 1: Lista di frequenza dei proverbi reperiti con approccio *corpus-driven* in itWaC

PROVERBI DA ITWAC	
Lista di frequenza	
PROVERBI DA ITWAC	OCCORRENZE
non c'è due senza tre	16
aiutati che Dio t'aiuta	13
l'unione fa la forza	13
chi ben comincia è a metà dell'opera	11
tra il dire e il fare c'è di mezzo il mare	11
una mela al giorno leva il medico di torno	11
chi va piano va sano e va lontano	10
il lupo perde il pelo ma non il vizio	9
non tutto il male vien per nuocere	9
chi lascia la via vecchia per la nuova, sa quel che lascia e non sa quel che trova	8
chi trova un amico, trova un tesoro	8
l'abito non fa il monaco	8
la gatta frettolosa fece i gattini ciechi	8
la mamma dei cretini è sempre incinta	8
sbagliando s'impara	8
chi rompe paga e i cocci sono suoi	7
chi si contenta gode	7
errare è umano, perseverare è diabolico	7
il buon giorno si vede dal mattino	7

l'occasione fa l'uomo ladro	7
morto un papa, se ne fa un altro	7
nessuno è profeta in patria	7
non c'è peggior sordo di chi non vuole sentire	7
anno nuovo vita nuova	6
dimmi con chi vai e ti dirò chi sei	6
due x sono meglio di uno	6
mal comune, mezzo gaudio	6
meglio soli che male accompagnati	6
moglie e buoi dei paesi tuoi	6
paese che vai usanza che trovi	6
una rondine non fa primavera	6
a caval donato non si guarda in bocca	5
can che abbaia non morde	5
chi dorme non piglia pesci	5
chi fa da sé fa per tre	5
chi non risica, non rosica	5
chi troppo vuole nulla stringe	5
di buone intenzioni è lastricato l'inferno	5
fatta la legge trovato l'inganno	5
fidarsi è bene, ma non fidarsi è meglio	5
gallina vecchia fa buon brodo	5
il meglio è nemico del bene	5
l'appetito vien mangiando	5
meglio tardi che mai	5
Natale con i tuoi, a Pasqua con chi vuoi	5
occhio non vede cuore non duole	5
quando il gatto non c'è i topi ballano	5
se la montagna non va da Maometto, Maometto va alla montagna	5
squadra che vince non si cambia	5

chi tace acconsente	5
a pensar male si fa peccato ma si indovina	4
al peggio non c'è fine	4
chi di spada ferisce di spada perisce	4
chi di speranza vive disperato muore	4
chi più spende, meno spende	4
il diavolo fa le pentole ma non i coperchi	4
impara l'arte e mettila da parte	4
l'ozio è il padre dei vizi	4
non si può avere la botte piena e la moglie ubriaca	4
non svegliare il can che dorme	4
non vendere la pelle dell'orso prima di averlo preso	4
patti chiari, amicizia lunga	4
prevenire è meglio che curare	4
se son rose fioriranno	4
sposa bagnata sposa fortunata	4
tra moglie e marito non mettere il dito	4
chi bello vuole apparire un poco deve soffrire	3
chi cerca trova	3
chi è causa del suo mal pianga se stesso	3
chi la dura la vince	3
chi la fa l'aspetti	3
dagli amici mi guardi iddio, ché dai nemici mi guardo io	3
fa più rumore un albero che cade di una foresta che cresce	3
il fine giustifica i mezzi	3
il riso fa buon sangue	3
il silenzio è d'oro	3
l'epifania tutte le feste si porta via	3
l'occhio del padrone ingrassa il cavallo	3
la curiosità uccise il gatto	3

la fortuna aiuta gli audaci	3
la speranza è l'ultima a morire	3
le disgrazie non vengono mai sole	3
lontano dagli occhi lontano dal cuore	3
non c'è fumo senza arrosto	3
non si getta via l'acqua sporca col bambino dentro	3
ride bene chi ride ultimo	3
sono sempre i migliori quelli che se ne vanno	3
tanto va la gatta al lardo che ci lascia lo zampino	3
tutto è bene quel che finisce bene	3
val più la pratica che la grammatica	3
a buon intenditor poche parole	2
a nemico che fugge ponti d'oro	2
a san martino ogni mosto è vino	2
agosto, moglie mia non ti conosco	2
al cuor non si comanda	2
anarchici a vent' anni conservatori a trenta	2
anche l'occhio vuol la sua parte	2
anno bisesto anno funesto	2
cercare (in un sacco di noci ci sta ancora un sacco di riso)	2
chi ha denti non ha pane e chi ha pane non ha denti	2
chi muore giace, chi vive si dà pace	2
chi paga compra	2
chi primo arriva meglio alloggia	2
chi semina vento raccoglie tempesta	2
chi si loda s'imbroda	2
chi va al mulino s'infarina	2
chi zappa la terra semina le fave!	2
con la morte si estingue l'inimicizia	2

contro la stupidità umana neanche gli dei possono niente	2
cuor contento il ciel l'aiuta	2
del senno di poi son piene le fosse	2
dio li fa e poi li accoppia	2
gli amici non permettono agli amici di guidare ubriachi	2
gli assenti hanno sempre torto	2
gli ultimi arrivati saranno i primi	2
il bove dice cornuto all'asino	2
il tempo è galantuomo	2
il troppo storpia	2
il vino fa buon sangue	2
in amore e in guerra tutto è permesso (Lapucci)	2
in mancanza di cavalli ... si corre con gli asini	2
l'amicizia raddoppia la gioia e dimezza il dolore	2
l'amico si conosce nel bisogno	2
l'amore non è bello se non è litigarello	2
l'erba del vicino è sempre più verde	2
l'olivo benedetto vuol trovare pulito e netto	2
la buona conoscenza è la base per la buona amicizia	2
la curiosità è femmina	2
la farina del diavolo va tutta in crusca	2
la fortuna è cieca	2
la vendetta è un piatto che va gustato freddo	2
le bugie hanno le gambe corte	2
meglio un morto in casa che un marchigiano fuori dalla porta	2
meglio un uovo oggi che una gallina domani	2
mens sana in corpore sano	2
mentre i medici discutono il malato muore	2
mors tua vita mea	2

nella botte piccola c'è il vino buono	2
nessuna nuova buona nuova	2
non è bello quel che è bello ma è bello quel che piace	2
non è mai tardi per imparare	2
non esiste una buona meta per chi non sa dove va	2
non rimandare a domani quello che puoi fare oggi	2
non si può andare in paradiso a dispetto dei santi	2
non si può fare d'ogni erba un fascio	2
non si vive di solo pane	2
non vedono la foresta a causa degli alberi	2
nulla di nuovo sotto il sole	2
ogni medaglia ha il suo rovescio	2
partire è un po' morire	2
piove governo ladro	2
piove sul bagnato	2
quando si ha un martello in testa si vedono tutti i problemi sotto forma di chiodi	2
sacco vuoto non sta in piedi	2
scherza coi fanti ma lascia stare i santi	2
si fa fuoco con la legna di cui si dispone	2
tanto fumo e poco arrosto	2
tanto tuonò che piovve	2
tira più un pelo di f che un carro di buoi	2
tra i due litiganti il terzo gode	2
tutto il mondo è paese	2
una mano lava l'altra e tutt'e due lavano il viso	2
casa dolce casa	2
[i fatti che] sono argomenti dalla testa dura	1
30 giorni ha novembre con aprile * giugno * e settembre	1
a carnevale ... ogni scherzo vale!	1

a chi nasce sfortunato piove sul didietro anche se sta seduto	1
a chi tutto a chi niente	1
a forza di tenere aperta la bocca si sono chiuse le orecchie	1
a giovane eremita vecchio diavolo	1
a mali estremi estremi rimedi	1
a puttaparthi non si muove foglia che sai baba non voglia...	1
a sant' anna il 26 di luglio il lago si prenda sette vite	1
a voler troppo fare il bassotto si rischia di rimanere un tappeto	1
abbiate pietà di ben haddou	1
acca ' nisciuno e' fesso!	1
acqua passata non macina più	1
affrontare la vita come fa un tappeto di camomilla che più è calpestato e più si propaga	1
aggiungendo fuoco a fuoco	1
al lavoro con chi vuoi ma nelle vacanze con i tuoi	1
al meglio ci si abitua subito	1
all' annunziata ogni verme si risveglia	1
alla luna di settembre sette altre lune s'inclinano	1
alle parole seguiranno i fatti	1
amici da raccomandare a dio che dai nemici mi guardo io	1
amico con tutti fedele alle mie idee	1
amor senza baruffa fa la muffa	1
anche i viaggi di mille leghe iniziano dal primo passo	1

anche se esso dice e fa sempre le stesse cose non si ripete mai	1
aprile dolce dormire	1
aspetta cavallin che l'erba cresca?	1
attira più una goccia di miele che un barile di aceto	1
bacco tabacco e venere riducono l'uomo in cenere	1
balzano da due più forte di un bue	1
battere il ferro finché è caldo	1
beati i primi se gli ultimi saranno onesti	1
bella l'insegna buona la locanda	1
belle o brutte si sposan tutte	1
buon prete si fa frate e cattivo frate si fa prete	1
c'è chi sale e c'è chi scende e c'è pure... chi in c se la prende!	1
cadere dalla padella nella brace	1
canavesano coltello in mano	1
cancelli di santa chiara	1
cane non mangia cane...	1
casa è dove risiede il tuo cuore'	1
cercare il famoso ago nel pagliaio	1
cerchio lontano pioggia vicina e viceversa	1
che 'senigallia e' mezza ebrea e mezza canaglia	1
che chi indica la luna e c'è chi guarda il dito	1
che cipolle sono se non fanno piangere?	1
che cosa sognano le oche? il granturco	1
che il gioco non vale la candela	1
che ride bene chi ride ultimo	1
che se maometto non è profeta in patria allora la montagna va piano e va lontana	1

che tutto avviene gradualmente e chi troppo vuole niente ha	1
chi abbandona la partita la vince	1
chi ama vive di più e meglio	1
chi corre prima o poi inciampa	1
chi crede al ghiaccio/finisce all'addiaccio /niente è eterno/neppure l'inverno	1
chi crede e ama il cristo è come la rosa alla quale nessuno può impedire che il suo profumo si diffonda per tutti	1
chi cuce indosso scava il fosso	1
chi di verde si veste di sua beltà si fida	1
chi disprezza vuol comprare!!!	1
chi entra papa in conclave ne esce cardinale	1
chi fa ... falla	1
chi fabbrica su suolo altrui perde la calce le pietre e la sabbia	1
chi ha fame non ha sonno	1
chi ha la salvia nell'orto ha la salute nel corpo	1
chi ha molto riceverà ancora di più e sarà nell'abbondanza; chi ha poco gli porteranno via anche il poco che ha	1
chi ha tempo non aspetti tempo	1
chi ha tempo non aspetti tempo	1
chi nasce mulo bisogna che tiri calci	1
chi non beve in compagnia è un ladro o una spia'	1
chi non fa non sbaglia	1
chi non ha coraggio non vada alla guerra	1
chi non usa le verghe odia suo figlio	1
chi non va avanti - torna inevitabilmente indietro	1
chi pecora si fa il lupo se lo mangia	1

chi perde denaro perde qualcosa chi perde l'onore perde molto chi perde la s perde tutto	1
chi pianta datteri difficilmente potrà vederli e mangiarli	1
chi più sporca la fa diventa priore	1
chi rimane scottato ha paura anche dell'acqua calda	1
chi risparmia spreca	1
chi sa è padrone di chi non sa	1
chi sa e parla non vive molto a lungo	1
chi sa fa e chi non sa insegna	1
chi sa fare fa chi non sa fare insegna	1
chi sbaglia paga	1
chi scambia il certo il pagamento dello straordinario per l'incerto la maturazione del premio variabile non si fa la fama di essere un volpone...	1
chi scorda il suo passato... avrà dei problemi con il futuro	1
chi si è scottato con l'acqua calda ha paura anche di quella fredda	1
chi si estrae dalla lotta è un gran fijo de na	1
chi si fa gli affari suoi campa cent' anni!	1
chi si vanta da solo vale un fagiolo	1
chi spende più tempo meno tempo spende	1
chi sta con lo zoppo impara a zoppicare	1
chi sveglia una tigre addormentata rischia la vita	1
chi va a roma perde la poltrona	1
chi va con lo zoppo impara a zoppicare	1
chi va con lo zoppo va sano e va lontano	1
chi va per certi mari certi pesci e certe chiglie piglia	1
chi vive sperando muore c ?	1
chi vuol vivere a lungo in buona salute mangi salvia a maggio	1

chiacchiere e tabacchiere il monte [dei pegni] non le accetta	1
chiodo scaccia chiodo	1
cielo a pecorelle acqua a catinelle?	1
ciò che passa il convento	1
coi regali della befana/ci si gioca una settimana	1
come sia pericoloso togliere le brache ad un highlander	1
come un vecchio proverbio: meglio soli che male accompagnati	1
comunque più si è meglio è	1
con i santi non si scherza	1
con il tempo e con la paglia maturano le sorbe e la canaglia	1
con la forza non maturano nemmeno le pere	1
confermano la regola	1
conoscere pistoia e non solo di maria stefania bardi tesi	1
contrariamente al proverbio	1
cosa fatta capo ha	1
cuore non vede occhio non duole	1
da non potersi permettere ... di soddisfare la sete o la fame se gli fossero mancati i soldi	1
da qual pulpito vien la predica?	1
dà un cavallo a chi ti dice la verità ne avrà bisogno per fuggire	1
dal granaio la zizzania scompare e il buon grano resta	1
dalla padella alla brace	1
dalla via di più tardi si arriva alla piazza di mai	1
de gustibus	1
dei disegni son piene le mura'	1
del resto il gioco vale la candela	1
dell'attacco la miglior difesa	1

di aprile ce n'è uno	1
di troppi cazzi non è mai morto nessuno	1
dietro un grande uomo c'è sempre una grande donna	1
digiunare con la febbre mangiare col raffreddore:	1
dio aiuta chi si alza presto	1
dio dammi ragione per poter sopportare quello che non posso cambiare coraggio per cambiare quello che posso e saggezza per poter distinguere uno dall'altro	1
dio ha dato una bocca e due orecchie allo scopo di ricordarci che dobbiamo saper ascoltare il doppio di quanto parliamo	1
dio sorride se apri una porta; ed è triste se alzi un muro	1
dio vuole che tu conoscerai tante persone sbagliate prima di conoshere quella giusta in modo che quando la finalmente conoscerai tu sappia di essere grato	1
dire a suocera perché nuora intenda	1
domus mea tibi patet	1
donna al volante pericolo costante	1
dormi con i cani ti svegli con le pulci	1
dove ci son campane ci son puttane	1
dove ci sono tanti soldi l'acqua va per aria	1
dove esiste una buona tradizione bisogna seguirla dove non esiste bisogna crearla	1
dove passa il treno non cresce l'erba	1
dove ti sei fatto l'inverno ti fai pure l'estate	1
due ebrei tre opinioni	1
dura di più tante volte un recipiente incrinato di uno sano	1
è caduto da cavallo e sostiene di esserne sceso	1
è caduto da cavallo e sostiene di esserne sceso	1
è inutile piangere sul latte versato	1

è proprio vero che il sazio non crede al digiuno	1
e se il tempo è denaro ... allora è possibile investirlo	1
e' difficile ... trovare un gatto nero in una stanza buia	1
egli ha salvato la capra e i verzi cavoli	1
erba che non ha radice muore presto	1
erba marcia che non muore mai	1
esca grande pesce grande [errore presente nel testo]	1
esistono anche coloro che predicano male e razzolano bene	1
facciamo meglio ciò che ci piace	1
fai quello che il prete dice ma non fare quello che il prete fa'?	1
fare l'indiano	1
fatte il nome e fregatenne	1
febbraietto corto e maledetto	1
febbraio é corto anche se é amaro	1
fritto è buono tutto	1
gallina che mangia sassi sa il culo che ha	1
gatta ci cova	1
ghiri che dormiranno pure di giorno ... ma di notte si tolgono la fame rosicchiando tutto	1
gioca il sole prima che albeggi	1
giugno ciliege in pugno	1
gli amici dei miei amici sono anche miei amici	1
gli amici? pochi ma buoni	1
gli irlandesi sono in pace solo quando lottan	1
gli occhi servono a poco se la mente è cieca	1
gork e mork mi guardino dagli amici che dei nemici mi guardo io	1

gratta il russo e troverai il tartaro	1
guarda ciò che sono e non da chi son nato	1
hai o l'una o l'altra qui bisogna avere tanta testa e buona gambe	1
hai voluto la bicicletta? e adesso pedala'	1
ho preso due fette di salame e me le sono messe sugli occhi	1
ho sempre cercato di fare da me quello che molti cercano di fare in tre	1
homo homini lupus! l'uomo è il lupo degli uomini	1
i buoi hanno le corna	1
i confetti non sono per i somari!!!!	1
i diavoli si nascondono negli specchi	1
i figli nascono dai genitori ma non per i genitori	1
i figli si baciano nel sonno	1
i litigi fanno diventare amici i nemici...	1
i maiali secchi sognano le ghiande!!!	1
i pensieri e i ragionamenti di qualcuno influenzano il suo carattere o il suo comportamento	1
il bianco copre cento bruttezze	1
il bisognino con quel che segue	1
il cane è geloso della propria coda	1
il cane è tornato al suo vomito e la scrofa lavata è tornata ad avvolgersi nel brago	1
il carbonaro a casa sua è padrone!	1
il diavolo è sempre meno brutto di come lo si dipinge	1
il figlio resta figlio finché si sposa ma la figlia resta figlia per tutta la vita"	1
il gioco è guerra e ha il diavolo nel core il perdere però fa cattivo sangue	1
il ino buono sta nelle botti piccole	1

il ladro non è solo chi ruba ma anche chi gli regge il sacco	1
il lavoro ti nobilita nel farlo	1
il loro regno si distingue in tre fasi: vendetta? costruzione colloquio con lo spirito santo	1
il mare prende ciò che gli viene tolto	1
il matrimonio è la tomba dall'amore	1
il mattino ha l'oro in bocca	1
il medico pietoso fa la piaga purulenta	1
il mondo è fatto a scale c'è chi scende e c'è chi sale...	1
il mondo e' bello x ' e' vario!!	1
il paese è piccolo e la gente mormora	1
il pesce grosso mangia il pesce piccolo	1
il pesce puzza dalla testa	1
il poco non ha mai fatto male a nessuno	1
il pretino della cura benedice ragni e spazzatura	1
il pretino della cura benedice ragni e spazzatura	1
il sazio non crede al digiuno	1
il summano ha il cappello; se oggi fa brutto tempo domani sarà bello	1
il tempo e denaro!	1
il tempo è rimasto da sposare perché vuol fare quello che vuole	1
il tutto è maggiore della somma delle sue parti	1
il vizio ... purtroppo resta	1
inghilterra o spagna purché se magna	1
intantoè meglio vivere un giorno da leone che cent'anni da pecora e poichi si fa pecora il lupo lo mangia	1
invecchiando si migliora...	1
invece che finire dalla padella nella brace	1
io gioco è bello qnd dura poco	1

l'acqua di aprile ogni goccia un barile	1
l'altezza e' mezza bellezza	1
l'amore è cieco...	1
l'amore rende ciechi riguardo al futuro e nel presente esclude tutto ciò che l'ingombra	1
l'apparenza inganna	1
l'ascensore va sempre rimandato indietro	1
l'elefante viene sconfitto dal topo non dal leone	1
l'esperienza insegna	1
l'essere/sta nell'avere	1
l'invidia è così magra e pallida perché morde e non mangia	1
l'ospite e' come il pesce	1
l'una tira l'altra	1
l'uomo propone dio dispone	1
l'uomo tra due dame fa la figura del salame	1
l'uomo vale per quanto sa fare	1
la bellezza è negli occhi di chi guarda	1
la casa nasconde ma non ruba'	1
la classe non è acqua	1
la corruzione dell'ottimo è pessima	1
la critica rock è gente che non sa scrivere che intervista gente che non ha nulla da dire per gente che non sa leggere	1
la distanza distrugge i falsi amori e rinsalda quelli veri	1
la donna con il petalo di una rosa	1
la familiarità conduce al disprezzo	1
la fede muove le montagne	1
la lingua inglese è quella che dà del voi alla moglie ai figli e agli amanti e dà del tu solo a dio	1
la madonna pedalava porcodio come filava	1

la migliore maniera di guardare il dente del lupo è quando quest'ultimo è morto	1
la notte porta consiglio'	1
la parte + dura da mozzicare è la coda?	1
la passione aguzza l'ingegno	1
la paura fa quaranta	1
la pazienza è la virtù dei forti	1
la propria libertà termina dove comincia quella altrui	1
la prudenza non è mai troppa!	1
la pulizia è una mezza virtù	1
la ragione sta nel mezzo	1
la scopa nuova pulisce meglio della vecchia	1
la somiglianza genera simpatia	1
la speranza è la virtù dei forti;	1
la squadra diretta da serniotti aveva fatto i conti senza il classico oste	1
la vecchia non voleva mai morire per imparare	1
la vera carità comincia da sé stessi	1
la verità viene sempre a galla	1
la voce del popolo leggi la maggioranza è la voce di dio	1
le cose frettolose non vengono bene	1
le cose ripetute piacciono	1
le donne non si sfiorino neanche con un fiore	1
le galline di linhò hanno tutte e sempre la coda storta	1
le parole che ...hanno fatto traboccare il vaso	1
le persone che condividono con gli altri sono raramente affamate	1
le strade sono ponti d'oro per i nemici	1
le tre grandi virtù del programmatore sono la pigrizia l'impazienza e la tracotanza	1

lo spada vien di notte con le scarpe tutte rotte...	1
lo stolto getta un anello nel lago e un milione di saggi non possono ripescarlo	1
ma non è tutto oro quel che luccica	1
ma non vedi la trave che c'e' nel tuo	1
ma torniamo un poco come dice il proverbio l'acqua al nostro molino	1
mai chiedere all'oste se è buono il suo vino?	1
mal voluto non è mai troppo	1
mala carne fa bono crossista	1
male non fare paura non avere	1
male non fare paura non avere	1
male non fare paura non avere	1
mangiare con gli ebrei e non con i cristiani	1
mangiare questa minestra	1
marzo che tinge mentre aprile dipinge	1
marzo ventoso e aprile piovoso portano un maggio bello e florido	1
mater semper certa pater incertus	1
mazza e panella fanno i figli belli pane e senza mazza fanno i figli pazzi...	1
meglio aver paura che buscarne	1
meglio il sapere che la ricchezza	1
meglio pochi ma buoni	1
meglio un asino vivo che un dottore morto	1
meglio una torta in 4 che una merda da soli	1
melius abundare quam deficere	1
mens sana in corpore sano	1
mi parlano e dimentico; mi fanno vedere e sentire e ricordo; mi fanno provare e capisco	1

molte cose vere vengono dette per scherzo	1
molte parole e pochi fatti ingannano i savì e i matti	1
molte perle siano destinate	1
né di venere né di marte	1
nè il bel tempo nè il cattivo tempo durano tutto il tempo	1
ne uccide più la gola che la spada	1
necessità fa virtù	1
nel dubbio astieniti!!	1
nella vita... succede di tutto e il contrario di tutto	1
nessuno è così povero da non avere nulla da dare agli altri e nessuno è così ricco da non avere bisogno degli altri?	1
nessuno nasce maestro	1
niente di nuovo sotto il sole	1
non affidare ai lupi la custodia del gregge	1
non c'è cibo di re più saporito del pane	1
non c'è futuro senza memoria	1
non c'è sabato senza sole non c'è donna senza amore	1
non c'è trippa per i gatti	1
non c'è un male che non possa rivelarsi anche un bene	1
non cade foglia che dio non voglia	1
non costa nulla regalare al prossimo un sorriso	1
non disdegnate ciò che non potrete mai avere!	1
non è necessario dire tutto quello che si pensa ma è necessario pensare tutto ciò che si dice	1
non è possibile conoscersi reciprocamente prima di aver consumato insieme il sale	1
non è tutto oro ciò che luccica	1
non fare agli altri quello che non vorresti fosse fatto a te	1
non farti pecora che il lupo ti mangia	1
non lodare il vino all'amico	1

non metter piede dove prima non hai messo gli occhi!	1
non mettere mai il carraro davanti ai buoi	1
non mordere la mano che ti nutre	1
non puoi con un uovo in mano volere sia l'uovo che la frittata	1
non sappia la sinistra ciò che fa punto	1
non serve un cervello per poter parlare come non serve giudizio per poter criticare	1
non si fanno le nozze con i fichi secchi:	1
non si muove foglia che dio non voglia	1
non si nasce imparati	1
non si può smetter di mangiare perché si ha paura di cucinare	1
non sparate sulla croce rossa	1
non tutte le ciambelle riescono col buco'	1
nuovo è bello	1
o detto in altri termini attento a ciò che desideri perché potresti ottenerlo	1
occhio non vede cuore non duole	1
odio il bambino di precoce saggezza	1
ogni giorno nasce un cucco	1
ogni popolo ha il governo che si merita	1
ogni uomo ha il suo prezzo	1
ognuno è artefice della propria fortuna del proprio destino	1
ognuno si cura il proprio orticello	1
ognuno soffre del proprio mestiere;	1
parla come mangi o scrivi come parli?	1
parla poco e ficca bene!!!	1
parlare a moglie perché suocera intenda	1
passata la festa gabbato il santo	1

pazzo e' colui che va a piedi per comodo del cavallo	1
peggio che andar di notte	1
pensare globalmente e agire localmente'	1
per onestà va detto e ribadito che è gigi l'inventor del baxeiquito	1
per saper comandare bisogna saper obbedire	1
pesce grande mangia pesce piccolo	1
pochi maledetti ma subito!!!	1
poco cacio fresco poco san francesco”	1
povera comitiva se il capofila e' cieco	1
povera comitiva se il capofila e' cieco	1
prendere due piccioni con una fava	1
prendila con filosofia	1
presta ad un tuo amico il tuo ultimo dollaro ma non prestargli mai la tua ascia a meno che tu sappia che la sa maneggiare e che non ti rovinerà il filo	1
prima bagnata ... e quindi fortunata	1
prima di tutto bisogna sentire le due campane	1
prima il dovere poi il piacere	1
prima pensa poi parla perché parole poco pensate portano pena più pentimento...	1
primavera in mare	1
prova tutto ed attieniti a ciò che è buono	1
provare per credere	1
proverbio sulla lana e la seta?	1
quando a grilli e quando a tordi	1
quando in un fosso hai un coccodrillo alle calcagna dimentica che il tuo scopo era bonificare la palude	1
quando la gatta non può giungere a mangiare il lardo afferma che non lo mangia perché puzza	1

quando la moglie non si trascura il marito non cerca l'avventura	1
quando la nave affonda i topi scappano'	1
quando la volpe non arriva all'uva allora l'uva è acerba	1
quando non puoi batterli unisciti a loro...	1
quando tuona vuol piovere	1
quando una tavola è imbandita con ogni ben di dio qualcosa finisce sempre col cadere e chi sta sotto la tavola può approfittarne	1
quattro occhi vedono sempre meglio di due	1
quelli di sinistra oggi sono più realisti del re	1
quello che non costa niente non vale niente	1
quello del prete va alla chiesa ndt piove sul bagnato	1
questi danni non vengono mai da soli;	1
rende difficile il facile attraverso l'inutile	1
rende difficile il facile attraverso l'inutile	1
rende più sano e fa andare più lontano	1
restare in piedi anche nel fango'	1
rispetta il cane per il padrone	1
roma ladrona sheva non perdona...	1
rompe le uova nel paniere	1
rosso di sera bel tempo si spera	1
santa lucia 12 dicembre e' il giorno piu' corto che ci sia	1
sapientis est mutare consilium	1
sarebbe meglio di niente	1
saremmo cornuti e bastonati!	1
scherzi con le mani scherzi da villano	1
scienza senza coscienza non è che rovina dell'anima	1
scodella vuota un solo problema; scodella piena tanti problemi	1
se bella vuoi apparire un poco devi soffrire	1

se hai dato dimenticatene se hai ricevuto ricordatene	1
se hai una testa da dieci euro còmprati un casco da dieci euro	1
se il bisogno è maggiore la salvezza è più vicina	1
se il fini giustifica i mezzi	1
se il nemico non puoi batterlo alleati	1
se il tuo vicino e' perseguitato che la tua tenda lo accolga e il tuo palmizio lo sfami	1
se in un pollaio ci sono anche solo cinque galline ... non ci sono di certo anche cinque galli	1
se non è il socio che va al cai il cai va dal socio	1
se non tolleri l'ingratitude non far del bene!	1
se pecora ti fai il lupo presto o tardi ti mangerà?	1
se un cane prende una certa abitudine come quella di addentare la bisaccia piena di cibo anche se lo scacci ritornerà alla carica	1
se volete lodare qualcuno fatelo in pubblico se volete lamentarvi di qualcuno fatelo in privato;	1
se vuoi il miele non ti puoi permettere di tirare calci all'alveare	1
secondo cui la virtù è più noiosa del peccato	1
semi semi il riso abbonda sulla bocca degli stolti	1
sempre dà di più l'avarò che chi è senza danaro	1
sempre in testa al mattino fuori dai denti al pomeriggio e sotto i piedi la sera	1
senza lilleri non si lallera	1
si cerca insomma di usare ... anche il bastone visto che la carota non basta	1
si chiude la stalla quando i buoi sono già scappati	1
si comincia incendiari e si finisce pompieri	1
si pensa a chi non c'e' che... ha sempre ragione	1

si perdona finchè si ama	1
si può affermare che la conoscenza di altre esperienze risulta nel campo del diritto del lavoro e delle relazioni industriali estremamente utile se non addirittura decisiva	1
si ricordi che tanto va la gatta al lardo - che ci lascia lo zampino	1
sia a dio che al demonio di lasciare le loro impronte nei dettagli della storia	1
sii te stesso o te stessa chi può essere meglio di te	1
similis cum similibus	1
soldi e amicizia tradiscono la giustizia	1
sole e acqua a volontà'	1
sole sugli ulivi acqua sui ciapp	1
sole sugli ulivi acqua sui ciapp	1
solo chi fa sbaglia	1
solo la follia è capace di prolungare la giovinezza	1
son nato che il mondo diceva morirò che il mondo dirà!	1
sono sempre gli stracci che volano	1
sotto la neve il pane	1
spesso anche l'uomo pazzo parla giudiziosamente	1
stai attento a quello che desideri - potresti ottenerlo	1
su internet [...] nessuno sa che sei un cane	1
su qui e su qua l'accento non va	1
successo non è un nome di dio	1
successo non è un nome di dio	1
sui se non si fonda la storia	1
tale il padre tale i figli	1
tale la vita tale la morte	1
tant ' è ladro chi ruba che chi para il sacco	1
tanti nemici quanti schiavi	1
tentar non nuoce	1

terra nera buon frumento	1
tieniti stretti i tuoi amici e ancora piu stretti i tuoi nemici...	1
timoroso della tua stessa ombra	1
tra tanti muli ci può stare un asino	1
tranquillo! il peggio deve ancora venire!	1
tutte le cose belle della vita uccidono o fanno ingrassare:	1
tutte le strade portano a roma	1
tutte le volpi al fine si rivedono in pellicceria	1
tutte mi vogliono e nessuno mi piglia...	1
tutti i gusti sono gusti	1
tutti i nodi vengono al pettine	1
tutto a roma ormai si fa all'ombra del berluscone non piu' der cuppolone	1
un albero che cade fa piu' rumore di una foresta che cresce	1
un bel tacer non fu mai scritto	1
un buon vicino di casa è meglio di un fratello lontano	1
un cammello è un mulo creato in team	1
un diavolo scaccia l'altro e tanto meglio se quest'altro è il peggiore	1
un disegno vale più di mille parole '	1
un gangster che fa una brutta fine dice che il crimine non paga	1
un inganno tira l'altro	1
un tempo l'intelligenza valeva più dell'oro	1
un tempo la matematica non era un' opinione -	1
un uomo infastidito dice al suo compagno: basta smetti di lamentarti non ce la faccio più e l'amico gli risponde: e tu smetti di infilare il tuo coltello nel mio fianco	1

un uomo senza pancia è come un cielo senza stelle...	1
un vecchio che muore è come una biblioteca che brucia	1
una bugia a fin di bene vale più di cento verità	1
una ciliegina ne tira un' altra	1
una cuoca al tg 2	1
una femmina non e' mai bambina	1
una mezza verità è spesso peggio di una bugia	1
una notte ... può portare consiglio:	1
unire l'utile al dilettevole	1
uomo avvisato mezzo salvato	1
vale un fico secco	1
vedere per credere	1
vedi napoli e poi muor '	1
veloce come un ghepardo e longevo come una tartaruga	1
vendi a borsa alta e compra a borsa bassa	1
vendi e pentiti	1
verba volant	1
viaggiare bene è meglio che arrivare	1
voce del popolo che è voce di dio	1
voti denari e santità: metà della metà	1
vox populi vox dei	1

ALLEGATO 2: Lista di frequenza dei proverbi reperiti con approccio *corpus-driven* in frWaC

PROVERBI DA FRWAC	
Lista di frequenza	
PROVERBI DA FRWAC	OCCORRENZE
loin des yeux, loin du cœur	14
c'est en forgeant qu'on devient forgeron	12
qui veut noyer son chien l'accuse de la rage	12
l'union fait la force	11
l'habit ne fait pas le moine	11
il vaut mieux prévenir que guérir	10
qui ne tente rien n'a rien	10
tout vient à point à qui sait attendre	10
qui se ressemble s'assemble	9
charité bien ordonnée commence par soi-même	8
il n'y a pas de fumée sans feu	8
l'argent ne fait pas le bonheur	8
nul n'est prophète en son pays	8
qui aime bien châtie bien	8
aide-toi le ciel t'aidera	7
jamais deux sans trois	7
mieux vaut tard que jamais	7
quand on aime on ne compte pas	7
qui sème le vent récolte la tempête	7
tel père tel fils	6
en mai fais ce qu'il te plait	6
la nuit porte conseil	6
qui trop embrasse mal étreint	6

tant va la cruche à l'eau qu'à la fin elle se casse	6
c'est dans les vieux pots qu'on trouve les bonnes soupes	5
celui qui n'a qu'un marteau considère tout problème comme un clou	3
en avril ne te découvre pas d'un fil	9
il ne faut pas vendre la peau de l'ours avant de l'avoir tué	5
le temps c'est de l'argent	5
les cordonniers sont toujours les plus mal chaussés	5
on n'est jamais si bien servi que par soi même	5
à chacun son métier et les vaches seront bien gardées	4
avoir le beurre l'argent du beurre et la crémière	4
dis-moi ce que tu manges, et je te dirai ce que tu es	4
divise afin de régner (Maloux)	4
l'enfer est pavé de bonnes intentions	4
la critique est aisée, et l'art est difficile	4
la peur donne des ailes	4
les chiens aboient, la caravane passe (Maloux)	4
les petits ruisseaux font les grandes rivières	4
plus on est de fous, plus on rit	4
quand le sage montre la lune l'idiot regarde le doigt	4
qui dort dîne	4
qui ne dit mot consent	4
qui va lentement va sûrement	5
qui vole un œuf vole un bœuf	4
tous les chemins mènent à rome	4
à quelque chose malheur est bon	3
chat échaudé craint l'eau froide	3
donne un poisson à un homme il aura à manger pour un jour apprends -lui à pêcher il aura à manger pour tous les jours de sa vie	3
en france tout finit par des chansons	3

l'amour rend aveugle le mariage rend la vue	3
l'avenir appartient à ceux qui se lèvent tôt	3
l'occasion fait le larron	3
la fin justifie les moyens	3
la parole est d'argent mais le silence est d'or	3
le coeur a ses raisons que la raison ne connaît point	3
le poisson pourrit par la tête	3
les écrits restent les paroles s'envolent	3
les promesses n'engagent que ceux qui y croient	3
on attrappe plus de mouches avec du miel qu'avec du vinaigre	3
on choisit ses amis on ne choisit pas sa famille	3
on ne fait pas d'omelettes sans casser d'œufs	3
plus c'est gros	3
pour souper avec le diable il faut une longue cuillère	3
qui n'aime pas les bêtes n'aime pas les gens	3
rien ne sert de courir, il faut partir à point	3
rira bien qui rira le dernier	3
toutes les bonnes choses ont une fin	3
à chaque jour suffit sa peine	2
à cheval donné il ne faut pas regarder la dent	2
a la saint valentin elle me caresse la main vivement la sainte brigitte qu'elle me caresse la	2
au pays des aveugles les borgnes sont rois	2
avant de relever la paille chez ton voisin regarde plutôt la poutre dans ton'il	2
avec des amis pareils on n'a pas besoin d'ennemis	2
avec des si on met paris en bouteille	2
ce sont les meilleurs qui s'en vont les premiers	2
comme on fait son plumard on se couche	2
faute de grives, on mange des merles	2

il faut battre le fer pendant qu'il est chaud	2
il faut qu'une porte soit ouverte ou fermée	2
il faut tourner sept fois sa langue dans sa bouche avant de parler	2
il n'est pire sourd que celui qui ne veut entendre	2
il n'est jamais trop tard pour bien faire	2
il n'y a que les imbéciles qui ne changent pas d'avis	2
il ne faut pas se fier aux apparences	2
il vaut mieux être seule que mal accompagné	2
l'un sème, l'autre récolte (Maloux)	2
la crainte [ou peur] du gendarme est le commencement de la sagesse	2
la liberté des autres s'arrête là ou la vôtre commence	2
la main qui pousse le berceau dirige le monde	2
le mieux est l'ennemi du bien	2
les absents ont toujours tort	2
mieux vaut être tête de souris que queue de lion	2
ne fais pas aux autres ce que tu ne voudrais pas que l'on te fasse	2
ne remets pas à demain ce que tu peux faire aujourd'hui	2
oeil pour oeil dent pour dent	2
on ne récolte que ce que l'on sème	2
peu importe le flacon pourvu qu'on ait l'ivresse	2
plus c'est long plus c'est bon	2
quand le chat n'est pas là, les souris dansent	2
qui cherche trouve	2
qui mange de la salade n'est jamais malade	2
qui se sert de l'épée périra par l'épée	2
qui veut voyager loin ménage sa monture	2
si jeunesse savait, si vieillesse pouvait !	2
tant qu'il y a de la vie il y a de l'espoir	2

tout ce qui brille n'est pas or	2
un homme averti en vaut deux	2
un malheur ne vient jamais seul	2
vous ne pouvez pas satisfaire tout le monde tout le temps	2
à la sainte catherine tout arbre prene racine	1
à la sainte luce les jours allongent d'un saut de puce	1
a rome fais comme les romains	1
avec le temps et la paille (...) les nèfles mûrissent	1
c'est au pied du mur qu'on voit le maçon	1
ce ne sont pas les recettes de cuisine qui manquent c'est la faim	1
ce que femme veut dieu le veut	1
ce qui est pénible dans une démocratie c'est le vote des électeurs	1
chassez le naturel il revient au galop	1
dans le doute abstiens -toi	1
dieu fait tout pour le mieux	1
donnez moi la force de changer ce qui ne peut etre accepter d'accepter ce qui ne peut etre changé et la sagesse de faire la distinction entre les deux	1
faire et défaire c'est toujours travailler	1
faite ce que je dis pas ce que je fais	1
février trop doux printemps en courroux	1
gants pourris point de sortie - gants troués ktm à l'arrêt	1
il n'est pas nécessaire de faire triste pour être sérieux ...	1
il n'y a pas d'amour il n'y a que des preuves d'amour	1
je parle pas anglais c'est normal je l'ai appris à l'école	1
l'essayer c'est l'adopter	1
la fete passee adieu le saint	1
la mesure de l'amour c'est d'aimer sans mesure	1
la neige au grenier le feu à la cave	1

la perdrix retourne toujours au nid ;	1
le bon programmeur n'est pas celui qui ne fait pas d'erreur c'est celui qui les prévoit	1
le grand père boit les petits enfants trinquent	1
le juste et l'équitable sont identiques et quoique tous deux soient désirables l'équitable est cependant préférable	1
lentement mais surement	0
les petits cadeaux entretiennent l'amitié et les grands cadeaux entretiennent l'amour	1
mieux vaut une petite kowa sur le terrain qu'une optolyth 100 dans un placard	1
neige en août fais l'plein d'mazout	1
nourris un chien il te mordra la main	1
on peut courtiser le diable mais il ne faut jamais se marier avec lui	1
passé le jour passée la fête	1
pinguin dans les champs hiver mechant	1
plus tu les fuient plus il te suivent plus tu les suient plus il te fuient	1
quand dans tatouage magazine tu passeras te la raconter tu pourras	1
quand le disciple est prêt le maître apparaît	1
quand on n'y connaît rien on s'abstient	1
quand on veut on peut	1
qui a santé il a tout ; qui n'a santé il n'a rien	1
qui veut etre belle au concours doit souffrir pendant l'entrainement	1
ronfleur du soir pêteur du matin	1
si tu marches comme un canard et si tu fais coin-coin comme un canard c'est que tu es un canard	1
si tu veux labourer droit accroche ta charrue à une étoile	1

si tu veux mentir parles du temps	1
sous un arbre de noyer gardes toi de te coucher	1
traduire c'est trahir	1
un de perdu dix de retrouvés	1
une chanson de betty c'est le bonheur qui sourit	1
2 coqs vivaient en paix une poule survint	1
53 en poudre j'y emmène ma soeur	1
à balayer devant votre porte	1
a beau mentir qui vient de loin	1
a chacun son étoile	1
a force d'aller à l'eau laalebasse se casse	1
a la saint-bernard pour moissonner il est bien tard	1
a le coeur sur la main	1
à nîmes il semble que les crocodiles - (...) - se mangent parfois entre eux	1
à piqûre d'abeille cérumen d'oreille	1
a trop regarder comment urinent les autres tu vas finir par te pisser dessus	1
a vouloir tout gagner on risque de tout perdre	1
après la soupe un coup de vin vole un écu au médecin	1
après le bonheur le malheur vient	1
araignée du matin entrain araignée du soir foutoir	1
aucun laboureur ne s'arrête pour la mort d'un homme	1
autres temps autres moeurs	1
avant d'avoir la couche encore faut -il avoir la sous-couche car l'une sans l'autre ne sert à rien et vice et versa	1
bien mal acquis ne profite jamais	1
bien souvent un simple proverbe vaut mieux que deux bons avis	1
bon pied bon oeil	1
bonne renommée vaut mieux que ceinture dorée	1

bûche bien brûlée flamme déclarée	1
c'est ceux qui en parlent le plus qui en font le moins	1
c'est de participer	1
c'est en se vautrant la tronche qu'on apprend à faire du vélo	1
c'est la goutte d'eau qui fait déborder le vase	1
c'est le ton qui fait la musique	1
c'est quand c'est dur qu'on reconnaît les durs!!	1
c'est toujours quand les gens ne sont plus là qu'on les regrette le plus	1
c'est toujours quand on a besoin de mouchoirs qu'on en a pas	1
c'était le calme avant la tempête	1
ça sert à rien de réinventer le roue	1
calomniez calomniez il en restera toujours quelque chose	1
car qui n'entend qu'une cloche n'entend qu'un son	1
ce n'est pas le matériel qui fait la photo	1
ce qui est dit n'est jamais définitivement dit	1
ce qui ne les empêche pourtant pas de passer c'est-à-dire d'avancer comme la caravane du proverbe	1
ce sont les questions que nous nous posons qui nous font avancer pas les réponses	1
celui qui agit doit subir	1
celui qui combat et s'enfuit pourra toujours se battre un autre jour	1
celui qui déplace les montagnes est celui qui commence à déplacer les petites pierres	1
celui qui est très haut en l'air peut tomber très bas	1
celui qui garde son corps garde une solide forteresse	1

celui qui ne sait pas et ne sait pas qu'il ne sait pas fuir - le celui qui ne sait pas et sait qu'il ne sait pas éduquer - le celui qui sait et ne sait pas qu'il sait éveiller - le celui qui sait et sait qu'il sait fuir - le	1
celui qui ne sait pas modifier sa pensée ne sera jamais capable de modifier la réalité	1
celui qui te donnera sa dernière tartine de Nutella sera ton ami	1
celui qui veut du miel il lui faut patienter à la piqûre des abeilles	1
celui qui a fait une chose peut la défaire	1
cent meuniers cent boulangers et cent tailleurs font trois cents voleurs	1
ceux qui récoltent les cocons ne sont pas autorisés à fumer	1
ceux qui sont ignorants du passé se condamnent par là même à le réinventer	1
ceux qui vivent dans des maisons de verre devraient s'abstenir de lancer des pierres	1
chacun voit midi à sa porte	1
chaque bête se fera suspendre par son propre pied	1
le charbonnier est maître chez soi	1
chassez le naturel il revient au galop	1
chemisette et bermuda	1
considérez vos rêves comme les enfants de votre âme et accordez-leur toute votre attention car ils sont à l'origine de vos accomplissements	1
contre mauvaise fortune faire bon cœur	1
cul vu n'est pas perdu	1
cymbidium au bureau ne fait que poireau	1
dans la rizière (...) le crocodile te guette grimpe sur la berge le tigre t'attend	1

dans la yourte bulle grave	1
de deux maux il faut choisir le moindre	1
de la discussion naît souvent la lumière	1
demander peut occasionner un moment de gêne ne pas demander peut donner une vie de gêne	1
depuis que fut noyée la ville d'ys / on n'en a point trouvé d'égale à paris	1
des goûts et des couleurs il ne faut pas discuter	1
dès qu'y a une connerie à faire	1
deux précautions valent mieux qu'une	1
dieu aime les pauvres et donne aux riches	1
dieux aide à qui se lève tôt	1
dites un mensonge une fois c'est un mensonge dites le dix fois c'est toujours un mensonge dites le cent fois il devient une vérité sur laquelle on peut asséoir une religion!!	1
donne - moi ta montre et puis plus tard je te donnerai l'heure contre un paiement bien sûr	1
donnez -moi le courage de changer ce que je peux changer la sérénité d'accepter ce que je ne peux pas changer et la sagesse de reconnaître la différence entre les deux	1
donnez un clou à un aragonais (...) et il l'enfoncera avec sa tête	1
elle [la chaîne des pyrénées] n'est visible que lorsqu'il va pleuvoir bientôt	1
elle tenait à vous le dire	1
en cas de grand danger le chemin du milieu mène droit à la mort	1
en février bon merle doit nicher	1
en france on n'a pas de pétrole mais on a des idées	1

en revenant de la chasse la nouvelle clio a trouvé une modus à sa place	1
en se droguant nous ne faisons que survoller le champ de coquelicot ; tandis qu'avec la spiritualité et plus exactement la méditation nous nous trouvons dans ce même champ de coquelicot (...)	1
épargne de bouche vaut rente de pré	1
épée et argent requièrent mains astucieuses	1
être désœuvré comme ali ben hassan	1
fâcheux à monter comme la montée saint-maurice ?	1
fait attention à tes souhaits ils pourraient se réaliser	1
familial personnel ou professionnel commence par soi - même	1
faut de tout pour faire un monde	1
faute de parler on meurt sans confession	1
femme varie	1
figuier et olivier ne meurent jamais sans héritier	1
fin mai début juin période d'examen	1
geekette devant l'eseo snobeuse à vélo	1
gel à l'étape en septembre bonne neige en décembre	1
grand diseu petit faiseu	1
grève à tout crin ça craint	1
grosse tête grosse migraine	1
heureux au jeu malheureux en amour	1
heureux l'élève qui tel la rivière suit son cours sans sortir de son lit	1
i l ne faut un jour jurer de rien	1
iI en va ainsi à Thay	1
ii faut que jeunesse se passe	1
il est facile d'apporter remède au mal quand le patient ignore son traitement	1

il est licite à un chacun et louable de combattre pour sa patrie	1
il fait bon vivre sous la crosse	1
il faut creuser les puits aujourd'hui pour les soifs de demain	1
il faut être deux pour réaliser un grand match	1
il faut être pris pour être appris	1
il faut manger pour vivre et non pas vivre pour manger	1
il faut que jeunesse se fasse	1
il faut que tout le monde vive	1
il faut savoir ménager sa monture donc son corps si on veut aller loin	1
il faut se méfier de l'eau qui dort	1
il faut se vouloir se battre sans vouloir gagner mais sans vouloir perdre	1
il faut trouver chaussure a son pied	1
il fo se kité souven pour aimer tjs	1
il n'est de bon bec que de paris	1
il n'est de pire sourd que celui qui ne veut entendre	1
il n'est pas de vent favorable à celui qui ne sait pas quel est son port de destination	1
il n'y a de chance que pour la canaille	1
il n'y a pas de bon vent pour celui qui ne sait pas où il va	1
il n'y a pas de problème qui n'ait pas sa solution surtout si elle est idiote	1
il n'y a pas meilleur messenger que soi	1
il n'y a rien de nouveau sous le soleil	1
il ne faut jamais avoir raison trop tôt	1
il ne faut pas juger un livre à sa couverture	1
il ne faut pas parler si tu veux pas avoir des problèmes	1

il ne faut point puiser aux ruisseaux quand on peut puiser à la source	1
il ne faut jamais dire jamais!	1
il vaut mieux avoir affaire à dieu qu'à ses saints!!!	1
il vaut mieux avoir les bourses en action que les actions en bourse	1
il vaut mieux être riche et en bonne santé que pauvre et malade	1
il vaut mieux investir dans une bonne soie que dans une bonne canne	1
il vaut mieux tuer le diable que si le diable vous tue	1
il vaut mieux vivre en concubinage et mourir mariés	1
il y a des honnêtes gens partout	1
il y a loin de la coupe aux lèvres	1
il y a trompe - l'oeil mais il n'y a pas trompe goût	1
ils sont capables du pire comme du meilleur mais dans tout les cas attendez vous à être surpris	1
inactif aujourd'hui radioactif demain	1
j'avais l'air (...) de celui qui cherche un cheval plein ou des oeufs de chameau	1
je crois ce que je vois et je vois ce que je veux	1
je déteste mon porte-monnaie lorsqu'il est vide...	1
je puis t'amener à la rivière mais je ne puis boire à ta place	1
je sens bien une troisième	1
jeu à barre de vie prend ton eau de vie pour la barre	1
jeu de main jeu de vilain	1
l'alcool tue tout	1
l'amitié finit parfois en amour ms l'amour rarement en amitié	1
l'amour est aveugle	1

l'amour vous mord vous cloue vous moud mais vous ouvre l'âme et le corps	1
l'arroseur arrosé	1
l'avaleur n'attend pas le nombre des âneries	1
l'eau courante ne peut croupir et le gond d'une porte n'est jamais verroulu	1
l'économie ne se décrète pas	1
l'ennemi de ton ennemi est ton ami	1
l'erreur est humaine	1
l'escalier se balaie par le haut	1
l'expérience c'est le nom qu'on donne à ses erreurs	1
l'expérience est une lanterne qui n'éclaire que celui qui la porte	1
l'experience n'attend pas le nombre d'année	1
l'expérience ne sert qu'à couvrir des crânes chauves	1
l'herbe est toujours plus verte ailleurs	1
l'histoire se répète	1
l'hiver n'est point rude saison qui fait rester à la maison	1
l'homme fait ce qu'il peut le destin fait le reste	1
l'homme propose et dieu dispose	1
l'homme qui est né pour être noyé ne sera jamais pendu	1
l'idiot qui sait se taire a l'air d'un sage ?	1
l'important c'est de participer	1
l'oisiveté est mère de tous les vices	1
l'on garde toujours le meilleur pour la fin	1
l'on n'invente rien de nouveau	1
l'un n'a plus que sa chemise (...) l'autre n'a plus rien	1
la bave du crapaud n'atteind pas la blanche colombe	1
la bave du crapeau n'atteind pas la blanche colombe	1
la beauté n'est pas dans l'objet observé mais dans l'oeil de l'observateur	1

la caque sent toujours le hareng	1
la chance sourit aux audacieux	1
la défaite c'est comme un éléphant on peut l'examiner par en bas en dessous ou sur le coté et (...) on le verra différemment	1
la djellaba ne fait pas le musulman	1
la faute en est aux dieux qui la firent si belle	1
la femme au volant c'est la mort au tournant	1
la femme de césar ne doit pas être soupçonnée	1
la femme vaniteuse ressemble à la grenouille parée de plumes de boeuf	1
la forêt ne renvoie jamais en écho que ce qu'on lui a crié	1
la glace a besoin de froid le feu de la chaleur moi de toi	1
la goutte vient de la goutte	1
la joie grandit en étant partagée	1
la lisière est pire que le drap	1
la main du parjure tremble lorsqu'il écrit	1
la meilleure manière de prendre les choses du bon côté c'est d'attendre qu'elles se retournent	1
la nature fait bien les choses	1
la oie pense que persons ne peu pas voir elle quend elle cache la tête en sous de neige	1
là où il y a à manger pour deux il y a à manger pour trois	1
la passion du foot pour un jour à créer le fc canebiere pour toujours lol	1
la patience est recompensée	1
la principale faille de sécurité ce n'est pas le materiel mais l'individu qui l'utilise	1
la prison est l'école du crime	1
la seule chose nécessaire au triomphe du mal c'est l'inaction des gens de bien	1

la soupe fait le soldat	1
la tasse est plus chaud que le soup qui est de dans	1
la valeur n'attend pas le nombre des années	1
la vengeance est un plat qui se mange froid	1
la verite sors de la bouche des enfants	1
la vie commence à quarante ans	1
la vie trouve toujours son chemin	1
la ville à un visage la campagne à une âme	1
le bon chemin est jalonné de sources	1
le caillou et le rocher coulent dans l'eau de la même façon	1
le champ du voisin est toujours plus vert	1
le chien mord , mais pas le renard	1
le cochon deja cuit n'a pas du tout peur du four	1
le crime ne paie pas mais le repentir apparemment si	1
le diable finit toujours par étrangler le sorcier	1
le diable ne sait jamais qu'il est le diable	1
le diable se cache dans les détails	1
le dromadaire voit toujours la bosse de son voisin	1
le fonio ne fait jamais honte à la cuisinière	1
le maçon en venant peut faire passer sa bourse sous la porte et s'en retourner	1
le mauvais temps c'est le temps qui ne change pas	1
le mensonge tient un moment la vérité tient toujours	1
le mérite perd de son éclat sans quelque récompense	1
le monde est petit	1
le petit déjeuner conditionne toute votre journée il doit donc être équilibré copieux et varié	1
le red bull c'est de l'or en barre!!	1
le renault fait du bon boulot le renard boulotte les canards	1
le sabotier est souvent le plus mal chaussé	1
le silence est d'or	1

le soin énergétique par les couleurs et les cristaux	1
le soleil accepte aussi de passer par les petites fenêtres	1
le soupir c'est le rabot de la vie	1
le temps fais l'oubli	1
le travail c'est la santé	1
le travail peut tout	1
les belles choses sont difficiles	1
les bonnes idées ne font pas le printemps	1
les bons comptes font les bons amis	1
les bons conseils pénètrent jusqu'au coeur du sage ; ils ne font que traverser l'oreille des méchants	1
les bons paient souvent pour les coupables	1
les chiens ne font pas des chats	1
les conseillers sont souvent les mauvais payeurs	1
les diamants sont éternels	1
les extrêmes se touchent	1
les génies ne meurent jamais	1
les grands diseurs ne sont pas les grands faiseurs	1
les grands imprudents dans ton cas un drain peut en cacher un autre	1
les hommes debouts encaissent et les faibles font des comptes	1
les hommes ressemblent plus à leur temps qu'à leurs pères	1
les jours passent et ne se ressemblent pas	1
les semaines passent et ne se ressemblent pas	1
les vacances c'est tout un travail de repos très difficile	1
les valets sont pires que les maîtres...	1
lions au dedans en plein air renards	1
lorsque le vin est tiré il faut le boire	1
loue la mer mais reste à terre	1

lui ferait lâcher l'ombre de l'elysée pour la proie du palais bourbon...	1
mais scénario bidon n'a qu'une parole	1
maintenant suis ton chemin	1
mais comparaison n'est pas raison	1
mais en parlant de lièvres...	1
mange on ne sait pas qui te mangera	1
manger comme un roi le matin comme un prince à midi et comme un pauvre le soir	1
marche ou crève	1
médisez	1
méfions -nous de nos amis nos ennemis je m'en charge...	1
méfions nous de l'eau qui dors...	1
météo pas nette restons sous la couette (ou à la buvette)	1
mettre la charrue avant les boeufs	1
mieux vaut être onze et tristes que trezéguet	1
mieux vaut être pauvre et bien portant que riche et malade	1
mieux vaut être riche et bien portant	1
mieux vaut perdre une minute de sa vie que sa vie en une minute	1
mieux vaut s'adresser à dieu qu'à ses saints	1
ne crachez pas dans la soupe qu'on vous a donné	1
ne fait pas à autrui ce que tu ne voudrais pas que l'on te fasse	1
ne jamais désespérer	1
ne jamais goûter le gâteau tant qu'il n'est pas complètement cuit	1
ne jamais négliger le coeur d'un champion	1
ne quitte jamais celui qui t'aime pour celui que tu aimes car celui que tu aimes te quittera pour celui qui l'aime	1
ne sait de quoi demain sera fait	1

ni ange ni bête	1
noël au balcon pâques au tison	1
nous n'avons que le berger que nous méritons	1
nous n'héritons pas de la terre de nos encêtres nous emprignons celle de nos enfants	1
nous Noël aux balcons et pâques aux tisons	1
nous sommes ce que nous mangeons	1
nul n'est irremplaçable les cimetières en sont pleins	1
nul n'est plus chanceux que celui qui croit à sa chance	1
nul n'est prophète en son pays	1
nul n'est sensé ignorer la loi	1
oignez vilain il vous poindra ; poignez vilain il vous oindra	1
on a le temps	1
on a toujours besoin d'un plus petit que soi	1
on attrape le menteur avant le boiteux	1
on dit que les extrêmes se rencontrent	1
on dit que si l'on veut perdre une amitié il ne faut jamais rien lui prêter	1
on entend bien le chêne qui fait beaucoup de bruit quand il tombe mais on entend pas le petite graine qui pousse et qui grandit	1
on laissât du temps au temp	1
on n'a rien trouvé de mieux que vieillir pour ne pas mourir jeune	1
on n'est jamais trahi que par ces amis	1
on n'y a jamais joué deux fois la même partie	1
on ne coud pas du drap neuf avec du vieux ça le déchire	1
on ne dit pas fontaine je ne boirai pas de ton eau	1
on ne fait pas boire un âne qui n'a pas soif	1
on ne joue pas avec le feu	1

on ne juge pas l'écriture d'un homme mais plutôt sa façon d'utiliser son stylo pour satisfaire ses fans	1
on ne peut courir deux lièvres à la fois	1
on ne peut pas être à la fois au four et au moulin	1
on ne peut pas être et avoir été	1
on ne prête qu'aux riches	1
on ne reveille pas un chat qui dort	1
on ne sort de l'ambiguïté qu'à ses dépens	1
on ne tire pas sur l'ambulance	1
on peut tromper tout le monde quelque temps on peut tromper un petit nombre tout le temps mais on ne peut pas tromper tout le monde tout le temps	1
on pourrait dire que deux préventions valent mieux qu'une	1
on reconnaît l'arbre à ses fruits	1
on réponds aux!!!!!!!! par un silence	1
on sait ce qu'on perd on ne sait pas ce qu'on gagne	1
ont fait du passé un ingrédient du présent des manifestants afin d'organiser au mieux ce présent	1
où sera le corps là se rassembleront les vautours	1
pâques au scanner Noël au cimetière	1
parfois le hasard fait bien les choses	1
parle peu et agit beaucoup	1
parmi les loups il faut hurler avec les loups	1
pas d'argent pas de Suisse	1
pas de bras pas de Chicoungououinia	1
pas de changements pas d'agréments	1
pas de nouvelles bonne nouvelle	1
penser avant de mourir c'est un peu comme crever avant de partir...	1

personne n'a jamais tout à fait tord même une horloge arrêtée donne l'heure juste deux fois par jour	1
plus ça change plus ça reste la même chose	1
plus haut tu vis plus vite tu vieillis	1
plus le mensonge est gros plus il passe	1
plus les choses changent plus elles restent pareilles	1
plus on pédale moins vite moins on avance plus davantage	1
porte-feuille plat n'implique pas avarice	1
poumpadour poumpa	1
pour insérer heureux insérons cachés	1
pour que la joie soit complète que la choucroute et la bière soient de la fête faut siffer	1
pour savoir ou l'on va il faut savoir d'ou l'on vient	1
pour vivre heureux vivons cachés	1
premier dans son village plutôt que deuxième à rome	1
qu'il n'est bonne pesche qu'en eaue trouble	1
quand bidochon sortir glacière soleil taper sur plage arrière	1
quand c'est bio dehors c'est bon dedans...	1
quand deux taureaux s'affrontent c'est l'herbe qui en fait les frais	1
quand il gèle bien fort il gèle à pierre fendre	1
quand il n'y a pas assez de foin à l'écurie les chevaux se bat-tent	1
quand l'argent parle la vérité se tait	1
quand la maison est faite la mort entre	1
quand le batiment va tout va	1
quand le foin manque au râtelier les chevaux se battent	1
quand les bulles explosent les pingouins se reposent	1
quand nous danserons à tulotef	1
quand on ne sait plus faire quelque chose on l'enseigne	1

quand on résiste si on n'a pas toujours raison on a bien souvent des raisons	1
quand on sait pas on se tait a bon entendeur!!	1
quand on tue le coq les poules s'entretuent	1
quand on voit ce qui existe et quand on voit ce qu'on se tape	1
quand tu ne sais pas où tu vas regarde d'où tu viens	1
quand tu pries bouges tes pieds	1
quand un arrive un autre part	1
quand un artiste rencontre un autre artiste ils se racontent leurs problèmes d'argent quand un banquier rencontre un autre banquier ils parlent d'art	1
quand un peuple n'ose plus défendre sa langue il est mûr pour l'esclavage	1
quand wall street éternue paris s'enrhume	1
quatre-vingt-dix-neuf moutons et un champenois font cent	1
que l'espoir fait vivre	1
que le jeu en valait la chandelle...	1
que peut faire un mort entre les mains de son laveur ?	1
que ta main droite ignore ce que donne ta main gauche	1
quelque éloigné qu'il puisse être le bon et franc lapin (...) meurt toujours dans son terrier	1
qui a de la sauge dans son jardin n'a pas besoin d'un médecin	1
qui a peur de l'enfer meurt de faim en hiver	1
qui ce moque du passé n'est pas digne du futur"	1
qui crache dans l'oreille d'un sourd vole une heure à dieu	1
qui d'espoir vit sésespéré meurs ?	1
qui n'entend qu'une cloche n'entend qu'un son	1
qui paie ses dettes s'enrichit	1
qui peut le moins peut le plus	1

qui prête aux amis perd au double	1
qui se justifie trop s'accuse	1
qui veut faire l'ange fait la bête	1
qui vient à montreuil verra la murène	1
qui voit ouessant voit son sang qui voit sein voit sa fin affirme le dicton	1
qui voit sein voit sa fin	1
quiconque entre dans le tête à goret en ressort toujours avec de la m aux pieds	1
rappelle toi qui t'a fait roi	1
rien dans les mains - même pas un sac - rien dans les poches	1
rouler en amassant mousse	1
s'il niche haut les beaux mois fondent en eau et s'il niche bas l'été sera sec et chaud	1
s'ils ne veulent pas risquer quelque chose en nombre il y aura toujours quelqu'un qui sera contraint de tout risquer	1
sans l'aide divine nous ne pouvons rien	1
seul celui qui marche sur la braise en ressent la douleur	1
seul le verre et l'or donnent une idée du prix de la sagesse	1
seule la victoire est belle	1
si je ne me lance pas des fleurs moi-même personne ne le fera	1
si le conquérant pacifique de la grande île n'est pas pour ses concitoyens le prophète méconnu	1
si tu as l'argent tu auras le pouvoir	1
si tu as un projet à un an plante une graine si tu as un projet à dix ans plante un arbre si tu as un projet pour la vie éduque tes enfants	1

si tu cognes la tête d'un technogratte contre une cruche et que ça sonne creux n'en déduis pas trop vite que la cruche est vide	1
si tu connaissais les vertus de l'artémise tu la porterais dedans ta chemise	1
si tu regardes le soleil en face l'ombre sera derrière toi	1
si tu veux déplacer les montages une petite cuillère et ta patience suffiront	1
si tu veux la meilleure place dans la maison du devras déplacer le chat	1
si tu veux la paix prépares la guerre	1
si tu vois un crapaud en hauteur c'est qu'il n'a pas la paix à terre	1
si un ami te trahit la première fois c'est de sa faute la seconde fois c'est de ta faute à bon entendeur	1
sois sage et tais -toi	1
solide comme un chêne ?	1
taper d'abord questionner après	1
témérité inconsidérée n'est pas courage	1
tout caporal a dans sa giberne le bâton de maréchal	1
tout ce qui entre fait ventre	1
tout ce qui est au greffe est perdu	1
tout ce qui est tatouer et a mouais	1
tout commence tranquillement après la pluie le soleil	1
tout est bien qui finit un jour	1
tout est tout est pour le mieux chez le meilleur des hommes	1
tout finit par s'arranger même mal	1
tout flateur vit au dépend de celui qui l'écoute	1
tout prend racine	1
tout travail mérite salaire	1

toute chose a un commencement	1
toute vie est contrôle	1
trois simples cordonniers valent un grand esprit	1
trop de macro des années 70 tue la macro des années 70	1
tu ne fais pas la paix avec tes amis tu fais la paix avec des ennemis	1
tu seras seul dans ton cercueil	1
ui fait de l'imagination la folle du logis	1
un bon bailleur fait bailler sept personnes	1
un bon inverti en vaut deux	1
un cheval une alouette	1
un chien mouillé n'en sèche pas un autre	1
un clou chasse l'autre	1
un cochon ne peut pas se permettre de dire que la cour du mouton est sale	1
un con debout ira toujours plus loin qu'un génie assis	1
un esprit sain dans un corps sain	1
un fou qui hurle fait plus de bruit que cent sages qui se taisent	1
un homme a besoin d'ellébore lorsqu'il a l'esprit troublé	1
un homme qui a mangé n'a plus de mémoire	1
un hylien perdu dans la forêt est un hylien de moins	1
un jour legionnaire toujours est bien réel	1
un linceul n'a pas de poche	1
un mauvais accommodement vaut mieux qu'un bon procès	1
un mauvais ouvrier a toujours de mauvais outils	1
un pape meurt on en fait un autre	1
un sceau d'eau jeté de temps en temps n'altère pas le rocher mais une goutte continuellement finit par le briser	1

un second mariage est un triomphe de l'espoir sur l'expérience; il faut ajouter qu'un premier mariage est une victoire de l'espoir sur la statistique	1
un skieur de rando est toujours à l'affût des bonnes conditions	1
une bonne grippe à nouvel an vous guérit de tous les maux pour le reste de l'année	1
une bouchée avalée n'a plus de goût	1
une haie entre voisins préserve l'amitié	1
une image vaut bien un long discours	1
une mauvaise herbe en tue trois bonnes et prend la place d'une quatrième ?	1
une opinion préconçue empêche la recherche de la vérité	1
une personne ne peut rendre tout le monde riche mais tout le monde peut rendre quelqu'un riche	1
une petite clef peut ouvrir une grosse porte	1
une tête bien faite est mieux qu'une tête bien pleine	1
une vidéo youtube vaut 1000 mots	1
ventre affamé n'a pas d'oreille	1
vieillir ce n'est pas ajouter des années de la vie mais ajouter de la vie aux années	1
vis comme si tu allais mourir demain! apprend comme si tu allais vivre éternellement...	1
vite entré encore plus vite sorti	1
voilà ça vient...	1
y ' aura t il des tourteaux sur votre prochain territoire	1