

# Tra lingue e risorse linguistiche: metodi di analisi contrastiva della semantica di verbi di azione

Caterina Cacioli

## 1. Introduzione

La metodologia di ricerca negli studi contrastivi, in particolare dell'ambito semantico, si compone di diverse strategie di raccolta e analisi dei dati. Le sfide individuate nella ricerca contrastiva in semantica risiedono, in prima istanza, nella scelta delle risorse usate per le analisi, e in seconda istanza nei termini di paragone e nella comparabilità dei dati. Negli ultimi decenni la linguistica contrastiva si è affermata con una proliferazione di lavori grazie anche a risorse adesso a disposizione della ricerca (EBELING 2016; HASSELGÅRD 2020), che hanno promosso nuovi studi e la complementazione di studi precedenti. Si tratta, in particolare, di *corpora* multilingue e paralleli, che mirano a procurare dati altamente comparabili. Ulteriori tecnologie sono però a disposizione e forniscono possibilità per estendere la prospettiva delle ricerche contrastive: oltre ai *corpora*, che rimangono lo strumento più usato (JOHANSSON 2007; EGAN 2013; GRANGER E LEFER 2020), le analisi statistiche avanzate stanno diventando uno standard anche nella ricerca contrastiva (ENGHELS 2020, per esempi: GRIES 2013; LEVSHINA 2015; GRIES, JANSEGGERS E MIGLIO 2020; CACIOLI E VERNILLO 2023). Permane invece una scarsità di lavori che usano risorse multimodali, componente che nella ricerca in semantica risulta di fondamentale importanza per integrare le relazioni tra forme, significati e referenti nel mondo. Lo scopo di questo articolo è proporre l'uso combinato di due risorse linguistiche (una multimodale, *Imagact*, l'altra testuale) per la ricerca

contrastiva nella semantica dei verbi di azione, in relazione a questioni metodologiche, alla selezione dei termini per la comparazione e alla comparabilità dei dati. Da una prospettiva semantica, una questione che merita attenzione è l'individuazione dell'oggetto della ricerca da un punto di vista concettuale (= selezionare un dominio semantico) o linguistico (es., la categoria *verbi* o *aggettivi*), senza sovrapporre questi due livelli. In questo lavoro si affronterà questa questione all'interno degli studi contrastivi e attraverso l'illustrazione dei vantaggi della risorsa Imagact e della combinazione delle risorse, descrivendo il metodo proposto attraverso esempi di un caso di studio sui verbi d'azione.

## 2. Sfide metodologiche delle analisi contrastive

La linguistica contrastiva, una disciplina di ricerca inaugurata negli anni '50 (VINOGRADOV 2023; NORDRUM, EBELING E HASSELGÅRD 2016), si concentra sull'analisi dettagliata e sottile di somiglianze e differenze tra due lingue, o un piccolo gruppo di lingue, vicine dal punto di vista linguistico o socioculturale (GAST 2012). Le analisi contrastive sono indirizzate su un aspetto specifico, a qualsiasi livello di analisi, e la loro specificità fa emergere aspetti caratterizzanti le lingue in considerazione che solitamente sfuggono alle generalizzazioni tipologiche (KÖNIG 2012; ENGHELS, DEFRANCO E JANSEGGERS 2020) o all'osservazione di tali lingue in isolamento. Il vantaggio dell'uso di due (o poche più) lingue è anche di potersi concentrare su un'analisi qualitativa, a cui ricerche di larga scala non possono arrivare. Le due prospettive, contrastiva e tipologica, sono però collegate: se gli obiettivi delle analisi contrastive sono di scoprire cosa accomuna o differenzia due lingue su un aspetto specifico, al contrario di studi tipologici che ricercano schemi generali, è anche vero che gli studi contrastivi aiutano a indirizzare gli studi tipologici e a dare una prospettiva più specifica nella ricerca delle categorie universali (KÖNIG 2012). Gli studi contrastivi, infatti, non sono di per sé generalizzabili. Le somiglianze o differenze negli studi contrastivi sono definite internamente a un sistema, linguistico o concettuale, in riferimento alle lingue contrastate. L'aggiunta di

una lingua e la revisione dei parametri di paragone genererà risultati diversi oppure emergeranno nuovi aspetti da valutare. L'indagine è quindi *locale* rispetto alle lingue scelte, ma con un vantaggio di personalizzazione delle categorie da comparare: se in uno studio di larga scala si sceglie una categoria linguistica per la comparazione, che dovrà essere un'accomodazione generalizzabile a tutte le lingue (si veda HASPELMATH 2010 per una panoramica sulle categorie linguistiche e i concetti comparabili), negli studi contrastivi c'è un più alto margine di adattabilità al fenomeno che si sta studiando. Somiglianze e differenze sono quindi l'aspetto fondamentale su cui riflettere nella pianificazione e revisione delle ricerche contrastive, e ne rappresentano due grandi sfide legate tra loro: la prima, è relativa alla selezione dei termini di comparazione, la seconda all'effettiva comparabilità dei dati scelti.

### 2.1 *Tertium Comparationis* e semantica

Il termine di paragone comune alle lingue studiate, cosiddetto *tertium comparationis* (KRZESZOWSKI 1990) è la componente vitale di ogni ricerca contrastiva (CONNOR & MORENO 2005). Al fine di confrontare le strutture linguistiche, le forme o i significati tra le lingue, si deve trovare un riferimento comune, una categoria sovraordinata su cui confrontare il fenomeno. Infatti, a seconda del termine di paragone, due oggetti possono risultare simili o dissimili, e stabilire il punto comune nei termini di comparazione influenza quindi l'intera analisi di un fenomeno (JOHANSSON 2007; ENGHELS 2020). La ricerca di un *tertium comparationis* guida diversi aspetti della programmazione delle analisi contrastive: nel caso degli studi di semantica, un punto focale è la definizione dei concetti su cui concentrare la ricerca, unitamente all'osservazione delle strategie linguistiche usate per istanziare questi concetti e, di conseguenza, quali risorse sono utili per l'analisi e rendere operativa la domanda di ricerca (ENGHELS ET AL. 2020): per esempio, selezionare un concetto come "oggetto su cui sedersi" o delle realizzazioni linguistiche (sedia, poltrona, sgabello...) in più lingue, avranno come esito una diversa selezione di termini di comparazione e dei metodi

di raccolta dati (così come diversi risultati) considerato che le lingue hanno una varietà di categorizzazione linguistica nella denominazione degli artefatti (MALT 2010).

In semantica contrastiva, i termini di paragone sono ricercati in primis nelle risorse lessicografiche, nell'etimologia o negli approcci traduttologici: un'etimologia in comune (ad esempio, COUSSÉ, VAN DER AUWERA 2012), definizioni nei dizionari (bilingue o monolingue) o equivalenze di traduzione. Riguardo la selezione etimologica, studi hanno evidenziato quanto lingue geneticamente vicine possano differire nell'espressione dei domini semantici (MAJID, JORDAN E DUNN 2015): per esempio, in MAJID ET AL. 2007 si dimostra che verbi che in lingue germaniche hanno un'origine comune (inglese *break* e svedese *bräcka*) non sono equivalenti semantici. Già KRZESZOWSKI 1990 aveva mostrato come l'uso degli equivalenti di traduzione sia erroneo in principio per la ricerca in semantica: per esempio, l'uso di *corpora* paralleli per l'estrazione di traduttori ricorrenti non assicura equivalenza semantica, dal momento che una buona traduzione non richiede necessariamente equivalenza semantica.

Un approccio strutturato agli studi di semantica contrastiva emerge in GODDARD E WIERZBICKA 2009: qui si descrive una metodologia contrastiva sulla semantica dei verbi che parte dalla creazione di una lista più esaustiva possibile di qualità che definiscono la semantica di un verbo, analizzando gli usi del verbo nel loro contesto di occorrenza, da testare con giudizi di competenza di parlanti nativi. Il *tertium comparationis* è un dominio concettuale, quello delle *physical activities* in verbi come gli inglesi *cut* e *chop* (italiano *tagliare*), dove la selezione del dominio è operata tramite forme verbali, partendo dai verbi inglese e selezionando i traduttori in polacco e giapponese. Questo procedimento può essere considerato non obiettivo perché sovrappone il dominio concettuale a quello delle forme linguistiche, senza un chiaro termine di riferimento del dominio concettuale stesso. Un altro metodo che si riscontra nella letteratura è lo studio contrastivo di elicitazioni dirette sulla base di un supporto visivo (immagini o video) che limita il dominio concettuale selezionato (tra gli altri, MAJID, BOSTER E BOWERMAN 2008). In questo caso, ci sarà una forte comparabilità delle

risposte, ma l'elicitazione non prenderà in considerazione la varietà di uso in contesti di occorrenza naturale o l'uso in contesti non rappresentati nei supporti visivi.

## 2.2 Comparabilità dei dati

Una volta stabiliti i termini e le modalità di comparazione, è necessario valutare i dati presi in considerazione. La comparabilità dei *corpora* è sicuramente minore, rispetto a quella del dato elicitato, che può essere tenuto più sotto controllo dal ricercatore, ma i *corpora* rimangono necessari per raccogliere un'ampia prospettiva sui contesti di occorrenza e gli usi. Agli albori della disciplina, i dati più in uso derivavano da grammatiche o risorse lessicografiche con conseguente limitatezza dei dati più spontanei a disposizione e un ampio uso dell'introspezione come metodo di indagine (ENGHELS 2020). L'avvento dei grandi database testuali dagli anni '80 ha dato una svolta alla linguistica contrastiva e i *corpora* sono diventati lo strumento più in uso in questo campo (come testimoniato dai numerosi volumi dedicati, tra gli altri: GRANGER, LEROT, AND PETCH-TYSON 2003; JOHANSSON 2007), dal momento che coprono una grande varietà di generi testuali e di lingue. Per le ricerche contrastive si fa riferimento a *corpora* comparabili o *corpora* paralleli, sulla base degli obiettivi della propria ricerca, tenendo a mente che la piena comparabilità di testi raccolti in un *corpus* provenienti da lingue diverse non potrà essere piena. I *corpora* paralleli, in largo uso negli studi traduttologici (NORDRUM, EBELING E HASSELGÅRD 2016;) e nell'addestramento di risorse per la traduzione automatizzata, sono composti da testi tradotti in due o più lingue: dai *corpora* bilingue e bidirezionali (ovvero, con traduzioni e originali di entrambe le lingue) a *corpora* di lingue europee (EuroParl, con 11 lingue, KOHEN 2005). La comparabilità dei *corpora* paralleli è piena per il tipo di testi che si confrontano (stessi testi, con le loro traduzioni), ma riportano effetti di traduzione (del cosiddetto *traduttorese*, FUSCO 2006) che possono rendere falsificabile ed erronea la comparazione. VOLANSKY, ORDAN E WINTNER 2015 hanno infatti dimostrato che le caratteristiche del traduttorese sono

individuabili con metodi di apprendimento automatico addestrati sui *corpora*, che riescono a distinguere testi originali da testi tradotti. Questi fattori rendono i *corpora* paralleli una risorsa parzialmente affidabile per studi contrastivi di semantica, non applicati alla traduzione. Sicuramente, le preferenze o ricorrenze di traduzione offrono un dato semantico interessante, tuttavia si trova un'alternativa nei *corpora* comparabili. Questi, contengono testi originali, assimilabili per caratteristiche linguistiche o extralinguistiche: il genere testuale, la competenza linguistica dello scrivente (nel caso di *corpora* di apprendenti L2), il dominio di appartenenza del testo o il metodo di raccolta. L'uso di *corpora* paralleli e comparabili combinati è auspicabile (MAURANEN 2005) per complementare le informazioni apportate da ognuna di queste risorse.

### 3. Combinazione di risorse

Come accennato nel § 2, la ricerca contrastiva si basa fortemente sui *corpora*. Tuttavia, con le nuove tecnologie, si incoraggia la combinazione di più risorse per ampliare le prospettive su un dato fenomeno e ottenere prove convergenti o divergenti dai risultati precedenti (MAURANEN 1999; GILQUIN 2000; ARPPE E JÄRVIKIVI 2007). Per questi motivi, si propone qui l'uso di un'ontologia multimodale e *corpora* comparabili per la ricerca contrastiva in semantica. Si prendono in considerazione esempi di verbi di azione, con particolare riferimento ad azioni su oggetti che provocano un cambiamento di stato irreversibile (ex. rompere un vaso), in tre lingue (italiano, inglese, svedese). L'applicazione della metodologia si inserisce nella letteratura di studi contrastivi su questo dominio concettuale (FILLMORE 1970; MAJID ET AL. 2007, VIBERG 2020) e sulla categoria linguistica dei verbi (VIBERG 2010, 2013; KOPECKA E NARASIMHAN 2012).

#### 3.1 Imagact: scene visive come *tertium comparationis*

IMAGACT è un'ontologia visuale dell'azione, multimodale e multilingua, che integra informazioni visive e linguistiche (MONEGLIA ET

AL. 2012; MONEGLIA ET AL. 2014). Contiene 1010 azioni concrete rappresentate sottoforma di brevi videoclip, mappati su 16 lingue. Ogni video è annotato con i verbi che una lingua può usare per descriverlo e una frase descrittiva standardizzata (soggetto-verbo-oggetto). La risorsa è liberamente consultabile ([www.imagact.it](http://www.imagact.it)) e viene estesa a nuove lingue tramite una procedura di annotazione con parlanti nativi della lingua target. Ci sono tre funzioni di visualizzazione delle informazioni. Il primo è l'uso di una *galleria* di azioni: i video sono divisi in 9 meta-categorie e l'utente può navigare tra le categorie (per esempio, modifiche a oggetti o espressioni facciali) e cercare le azioni di interesse secondo ispezione puramente visiva, per poi accedere all'informazione linguistica associata. La funzione *dizionario* consente all'utente di inserire un verbo in una data lingua in una barra di ricerca e scegliere una lingua target. La risorsa restituisce i video con cui è taggato il verbo ricercato e i verbi associati a quei video nella lingua target. Nella funzione *comparativa*, l'utente può inserire nella barra di ricerca due verbi (in una lingua per il confronto intralinguistico o in due lingue per un confronto interlinguistico). La risorsa restituisce una schermata di confronto tra video con cui sono taggati i verbi oggetto della ricerca, divisi in 3 colonne. Nella colonna di sinistra si trovano i video taggati con il verbo A, nella colonna centrale i video che condividono la possibilità di riferimento con entrambi i verbi e nella colonna di destra i video taggati con il verbo B. Questa risorsa offre un ampio quadro della varietà di azioni che ricorrono nella nostra vita quotidiana con un potente strumento di contrasto. La funzione contrastiva è infatti particolarmente interessante perché offre una possibilità di contrasto dettagliato tra lingue.

L'input di partenza per la scelta dei termini di comparazione può avvenire su base visuale o linguistica. Su base visuale, si seleziona una caratteristica che accomuna delle azioni tramite la galleria e si procede alla selezione delle scene. Si prosegue integrando la componente linguistica, scoprendo i verbi che le descrivono e li si compara. L'altra modalità è su base verbale, per esempio, selezionando due traducanti – due verbi che nei dizionari bilingue sono dati come traduzione reciproca –. La scelta di uno dei due metodi è dettata dalle proprie necessi-



In ognuna di queste scene, un oggetto perde la sua forma o parte della sua forma iniziale. Confrontando 3 lingue, vediamo la varietà di verbi a loro associati (tabella 1).

Italiano	Inglese	Svedese
rompere	break	göra/ta/ha sönder*
spaccare	smash	slå sönder
spezzare	snap	bryta, bryta av, bryta loss, bryta upp
spezzettare	open	Bräcka, bräcka upp
staccare	crack	falla sönder
sfasciare	crush	slita sönder
sfondare	damage	riva, riva sönder
schiacciare	bust	knäcka
aprire	break open	krossa
forzare	tear	spräcka
dividere		göra hål
sbattere		banka
danneggiare		dela
strappare		krocka
		förstöra
		öppna

\* Queste tre varianti sono semanticamente equivalenti in svedese, con un'alternanza regionale (VIBERG 2020). La particella sönder si traduce letteralmente con "a pezzi" ed è unita a *ta* (lett. prendere), *ha* (lett. avere, rendere), *göra* (lett. fare).

**Tabella 1** Verbi associati alle scene selezionate, nelle lingue italiano, inglese, svedese

Per prima cosa, possiamo valutare la selezione dei verbi. Dal punto di vista intra-linguistico, questo gruppo di verbi non sono necessariamente sinonimi o in relazioni lessicali (per esempio, di iperonimia). Essendo ogni video annotato con una varietà dei possibili verbi che possono riferirsi a quell'azione, la risultante annotazione restituisce un gruppo eterogeneo. Infatti, ogni azione può essere descritta facendo riferimento a diversi aspetti concettuali<sup>1</sup>: possiamo citare il caso di

<sup>1</sup> Per un approfondimento sulla questione, si veda il concetto di equivalenze locali in MONEGLIA 1993; MONEGLIA 2000.



*rompere-aprire*, nel contesto, molto frequente in cucina, della manipolazione di uovo al fine di accedere al suo contenuto (ex. Giulia rompe/apre un uovo) che selezionano un diverso aspetto dell'azione che si riferisce a una diversa area concettuale: da una parte l'accesso a un contenuto (*aprire*), dall'altra la manipolazione distruttiva dell'oggetto (*rompere*).

Da un punto di vista contrastivo, si nota una più ampia varietà di verbi svedesi rispetto all'italiano e all'inglese. Se consideriamo i verbi più generali che descrivono queste scene (e quindi, si applicano alla maggior parte delle scene), possiamo disegnare il nucleo linguistico delle scene: i verbi *rompere*, *break*, e *ha/ta/göra sönder* rispettivamente per l'italiano, l'inglese e lo svedese. Tutti e tre questi verbi delimitano eventi di rottura: ne deriva che un'ispezione visiva delle scene conduce a un raggruppamento piuttosto coerente di verbi e che i verbi nelle tre lingue si riferiscono a questa categoria concettuale con un certo grado di equivalenza.

A questo punto possiamo effettuare la ricerca su base linguistica. Per quanto riguarda questo dominio, la selezione visiva si sovrappone quasi completamente alla selezione linguistica effettuata ricercando i singoli verbi *rompere*, *break* e *ha sönder* – che abbiamo detto essere il nucleo centrale del dominio concettuale selezionato. Si aggiunge una sola scena: colpire lo schermo di un pc, che perde la schermata di funzionamento, riferibile a un evento di perdita di funzionalità ma non di integrità visibile esternamente (scena n. 26 nell'appendice). La situazione di sovrapposizione non si mantiene per altri verbi della selezione, che si estendono su altri domini concettuali. È il caso di verbi generali come *aprire/open/öppna* (rispettivamente, italiano-inglese-svedese) che solo in alcuni contesti di distruzione dell'oggetto al fine di accedere a un contenuto sono applicabili ai video selezionati, oppure del verbo *smash*, che ha un'estensione maggiore e condivisa in larga parte con *schacciare*. Lo studio contrastivo può passare alla valutazione delle associazioni dei verbi alle scene in queste lingue, per valutare la loro distribuzione. In svedese, le scene per *ta/ha/göra sönder* sono comprese nella stessa varietà di scene dell'italiano e dell'inglese, ma in minor numero. Si nota quindi che: a. c'è un'alta sovrapposizione dei verbi generali all'interno di un dominio selezionato visivamente, che indica che le tre lingue condividono una partizione concettuale delle azioni; b. *rom-*

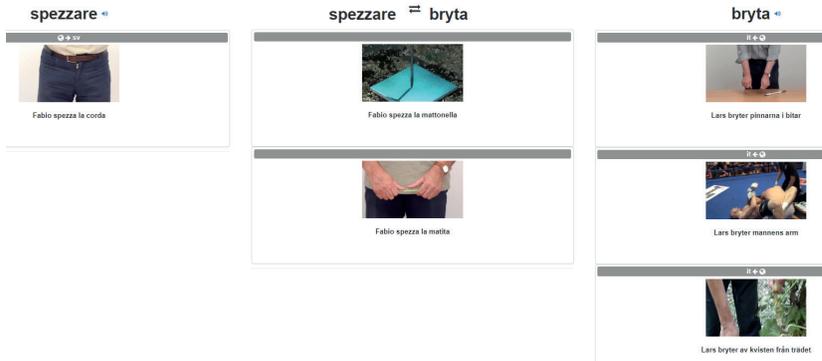


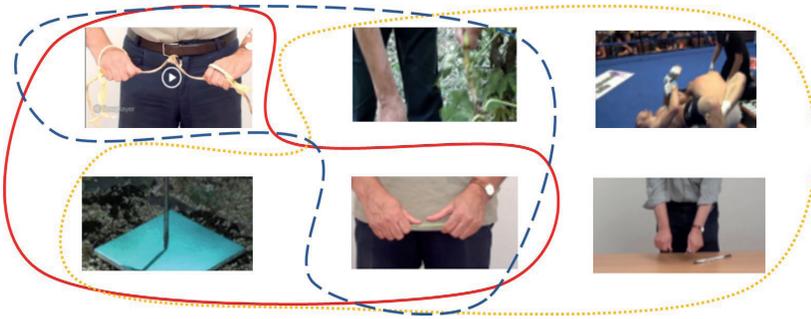
Figura 3 Esempio dell'interfaccia utente di Imagact nell'uso della funzione compare con i verbi *spezzare* e *bryta*

*pere* e *break* hanno una varietà di applicazioni alle scene più ampia di *ha sönder*, che per generalità alterna con *bryta*.

Passando al confronto degli altri verbi, più specifici, che descrivono queste scene, otteniamo ulteriori informazioni di natura contrastiva relativa alla partizione semantica del dominio concettuale. Con la funzione *compare* di Imagact, possiamo valutare i verbi a coppie, inserendo un verbo in una lingua A e uno in una lingua B, fino a esaurire le possibili combinazioni. La figura 3 riporta un esempio dell'interfaccia di Imagact e la figura 4 riporta in forma riassunta l'esito di tale comparazione per tre lingue.

Emerge che *spezzare-snap-bryta* riportano scene in comune, venendosi così a costituire un altro cluster semantico, sottogruppo della selezione iniziale. *Spezzare* e *bryta* condividono due scene, ma non le stesse di *snap* e *bryta*, e ancora una volta non le stesse di *spezzare* e *snap*. Non sono totalmente equivalenti ma catturano una porzione di realtà in comune. Tramite il confronto delle scene denotate da questi verbi si trae informazione contrastiva riportata in figura 4.

Un altro sottogruppo è composto dai verbi *spaccare-smash-slå sönder*. Qui notiamo che *smash* si estende a scene che non erano presenti nella selezione iniziale – ovvero scene in cui non c'è una distruzione ma un cambio nella forma dell'oggetto (corrispondenti all'italiano *schiaccia-*



**Figura 4** La linea continua delimita *spezzare*, la linea tratteggiata delimita *snap* e la linea puntata delimita *bryta*

re) –. Questo non accade con *spaccare* e *slå sönder*, che rimangono inclusi nella selezione di scene operata su base percettiva.

Alla luce dell'illustrazione della variazione dei verbi nel dominio semantico riportato sopra, vediamo che l'uso del mezzo visivo per disambiguare i concetti dalle forme linguistiche (e operare a un livello extralinguistico) è un grande vantaggio della risorsa, che consente di superare la questione relativa ai concetti e alle forme linguistiche nella ricerca in semantica. Le scene possono essere infatti usate in ottica contrastiva come *tertium comparationis*, il termine di riferimento comune su cui ricercare somiglianze e differenze che altrimenti non sarebbero esplicite. Infatti, questo tipo di informazione non è recuperabile nei dizionari o nei *corpora* (in cui mancando il mezzo visivo, l'azione non può essere giudicata). Esso fornisce una via d'uscita dalle implicazioni metalinguistiche che richiedono l'esplicitazione a mezzo linguistico del significato di un verbo e dal problema illustrato da GODDARD AND WIERZBICKA 2009 e all'alternativa delle descrizioni elicitate, nel § 2.

Sicuramente, l'estensione del *tertium comparationis* qui è limitata dalla natura stessa dei database che non possono contenere infinite scene visive. Catturare tutte le possibili azioni che possono essere compiute nella realtà è sicuramente impossibile. La risorsa è però

un'ampia e solida base di confronto tra verbi ai fini contrastivi, che può essere estesa ad altre lingue, mantenendo inalterato il riferimento di comparazione. Ciò permette inoltre di rendere meno *locali* le ricerche contrastive (come menzionato nel § 1).

Per uno studio contrastivo più completo sui verbi di azione, si deve però rendere conto di una dimensione importante: il contesto naturale di occorrenza e la varietà di contesto linguistico. Per queste ragioni, sarà importante combinare questa risorsa con i dati da *corpora*. Nel paragrafo successivo si suggerisce l'uso di *corpora* comparabili, presentando la famiglia di *corpora* TenTen, come strumento per esaminare la semantica del gruppo di verbi nei loro contesti naturali e privi di effetti di traduzione.

### 3.2 *Corpora* comparabili

I *corpora* qui proposti come valida risorsa per la linguistica contrastiva sono la famiglia di *corpora* TenTen (JAKUBÍČEK ET AL., 2013). Si tratta una collezione di *corpora* che conta ad oggi 43 lingue<sup>2</sup>.

Lo scopo di questa famiglia di *corpora* è raccogliere materiale autentico dal web selezionando i testi seguendo uno stesso protocollo per tutte le lingue. Viene usato uno strumento di *crawling*, ossia uno strumento di pulizia di testi che elimina il contenuto linguistico indesiderato (generalmente, quello non testuale, come i link e le intestazioni) per poi procedere con tokenizzazione, lemmatizzazione ed etichettatura dei generi testuali. Campioni di testi sono controllati poi con procedure manuali e automatiche per raffinare la selezione ed eliminare testi doppi (per esempio, uno stesso articolo pubblicato due volte). Per molte di queste lingue, nuove versioni dei *corpora* vengono rilasciate a vari intervalli temporali. Per esempio, l'italiano è stato rilasciato nel 2010, 2016 e 2020. Un preciso protocollo di creazione del *corpus* assicura che ogni lingua sia comparabile su determinati parametri, anche se l'aspetto che non può essere controllato è l'esatta comparabilità del contenuto dei te-

<sup>2</sup> Consultabili sulla piattaforma Sketchengine al link <https://www.sketchengine.eu/documentation/tenten-corpora/>.

sti – la componente culturale può infatti alterare la varietà e qualità dei temi trattati –. Basta pensare al peso che hanno le news nei contenuti web e l’impatto locale che gli argomenti delle news hanno sui contenuti. Questo aspetto, che può porsi come un limite alla comparabilità, va però ad aumentare la validità ecologica dei *corpora* in questione e di un’analisi contrastiva, che non è presente in *corpora* che comparano, per esempio, traduzioni di romanzi in varie lingue. In quel caso, ci sarà piena comparabilità dei temi trattati, ma minore naturalezza di occorrenza.

Il caso di studio qui mostrato utilizza tre *corpora* della famiglia Ten-Ten: ItTenTen 2016 per l’italiano, EnTenTen2015 per l’inglese e SvTen-Ten2014 per lo svedese e un campione di 500 occorrenze per verbi. Nella restante parte di questo paragrafo si evidenzieranno due aspetti: la dimensione concreto-astratto della semantica dei verbi e i temi/pazienti che vengono modificati, per esemplificare l’analisi contrastiva da combinare con Imagact.

L’informazione di Imagact è legata alle azioni concrete. Considerata l’astrattezza dei contenuti linguistici, saranno i *corpora* il luogo in cui cercare una misura della divisione tra riferimenti a concetti concreti o astratti per ogni verbo. La tabella 2 mostra i risultati dell’annotazione manuale<sup>3</sup> sulle occorrenze. I verbi riportati sono allineati nella tabella 2 secondo la loro vicinanza semantica, come discussa nel paragrafo precedente. Per lo svedese, si sono annotate due delle tre varianti regionali, *ha sönder* e *göra sönder*.

Italiano	Occorrenze concrete	Inglese	Occorrenze concrete	Svedese	Occorrenze concrete
Rompere	32%	Break	19.20%	Göra sönder Ha sönder	85.4% 89%
Spaccare	33.20%	Smash	60.20%	Slå sönder	60.60%
Spezzare	27.8%	Snap	16%	Bryta	14%

**Tabella 2** Percentuali di occorrenze che descrivono un’azione concrete per i tre

**3** L’annotazione manuale è stata condotta dall’autrice per le lingue inglese e svedese e da Paola Vernillo per la lingua italiana, con due annotatori esterni madrelingua, per ogni lingua, per la valutazione dell’accordo tra annotatori.

verbi generali delle tre lingue in osservazione in un campione di 500 occorrenze verbali produttive (escludi gli usi idiomatici) per verbo

Le occorrenze del campione mostrano una differenza netta nell'estensione di questi verbi. *Rompere* e *break* hanno un'alta componente metaforica, mentre il corrispettivo svedese si trova principalmente in contesti concreti (e simili percentuali di occorrenze concrete si ritrovano nelle due varianti svedesi *ha sönder* e *göra sönder*). L'italiano *spaccare* ha più usi metaforici rispetto ai vicini *smash* e *slå sönder*, mentre *spezzare*, mentre *snap* e *bryta* hanno più usi metaforici di *spezzare* (sebbene su percentuali inferiori).

Il ruolo tematico del paziente è significativo per l'indagine della semantica verbi di azione di modifica e manipolazione degli oggetti: le proprietà fisiche (come forma e dimensionalità) e funzionali (come l'oggetto viene usato e a quali fini) possono determinare le scelte verbali. Per il caso di studio qui discusso, l'annotazione di occorrenze mostra come i verbi svedesi *ha sönder* e *göra sönder* si estendono a oggetti di stoffa e vestiario in modo più sistematico rispetto all'italiano e all'inglese (1 occorrenza in *rompere*, nessuna in *break*, il 6% sia in *ha sönder* che *göra sönder*). Questo dato è rilevante dal momento che gli studi volti a indagare il concetto di *cambio di stato irreversibile in oggetti* rispettivamente ai verbi che lo realizzano (MAJID ET AL. 2007; MAJID, BOSTER E BOWERMAN 2008) hanno evidenziato, con task di elicitazione, come una dimensione principale della categorizzazione dei verbi d'azioni sia relativa ai materiali dell'artefatto e come la dimensione rigido-flessibile del materiale sia rilevante cross-linguisticamente nell'uso dei verbi, portando esempi di lingue in cui si fa una distinzione o meno (PYE 1996, lingua K'iche' Maya; BOWERMAN 2005, lingua inglese). Tramite *corpora*, si evidenziano sfumature nell'applicabilità di un verbo a categorie discrete (come rigido-flessibile) ed è possibile stimare il grado di estensione del fenomeno grazie alla varietà tematica ampia.

#### 4. Conclusioni

Gli studi di linguistica contrastiva, e in particolare semantica contrastiva, affrontano numerose sfide metodologiche. Per questo motivo le riflessioni sui metodi e la valutazione di strumenti sono ricorrenti e, di fatto, auspicabili per raffinare le strategie della ricerca contrastiva e produrre risultati più robusti. In questo contributo si è fornita una revisione di alcuni di questi aspetti (la comparabilità dei dati e i termini di riferimento) e si è proposto l'uso combinato di due risorse, *Imagact* e *corpora* multilingue comparabili, per contrastare lingue diverse. In particolare, si è mostrato come *Imagact* offra una soluzione alla ricerca del *tertium comparationis* per un'analisi contrastiva sui verbi azionali, utile a superare questioni di selezione del dominio concettuale o linguistico nella ricerca in semantica. Inoltre, la combinazione con i dati testuali è utile a fornire informazioni mancanti (es., sulla componente metaforica dei verbi) o supplementari (es., le associazioni verbo-argomenti).

#### 5. Bibliografia

- ARPPE-JÄRVIKIVI 2007 = ANTTI ARPPE, JUHANI JÄRVIKIVI, *Every Method Counts: Combining Corpus-Based and Experimental Evidence in the Study of Synonymy*, in «Corpus Linguistics and Linguistic Theory», 3 (2), 2007, pp. 131-59.
- BOWERMAN 2005 = MELISSA BOWERMAN, *Why Can't You 'Open' a Nut or 'Break' a Cooked Noodle? Learning Covert Object Categories in Action Word Meanings*, in *Building Object Categories in Developmental Time: 32nd Carnegie Symposium on Cognition*, a cura di L. Gershkoff-Stowe and D. Rakison, Mahwah, NJ: Lawrence Erlbaum, 2005, pp. 33-62.
- CONNOR-MORENO 2005 = ULLA CONNOR, ANA MORENO, *Tertium Comparationis: A Vital Component in Contrastive Rhetoric Research*, in *Directions in Applied Linguistics: Essays in Honor of Robert B. Kaplan*, a cura di Paul Bruthiaux, Dwight Atkinson, William G. Eggington, William Grabe e Vai-dehi Ramanathan, 2005.
- COUSSÉ-VAN DER AUWERA 2012 = EVIE COUSSÉ, JOHAN VAN DER AUWERA, *Human Impersonal Pronouns in Swedish and Dutch: A Contrastive Study of Man and Men*, in «Languages in Contrast», 12 (2), 2012, pp. 121-38.

- DEFRANCQ 2015 = BART DEFRANCQ, *Contrasting Contrastive Approaches*, in «Languages in Contrast», 15 (1), 2015, pp. 1-3.
- EBELING 2016 = JARLE EBELING, *Contrastive Linguistics in a New Key*, in «NJES Nordic Journal of English Studies», 15 (3), 2016, pp. 7-14.
- EGAN 2013 = THOMAS EGAN, *Tertia Comparationis in Multilingual Corpora*, in *Advances in Corpus-Based Contrastive Linguistics: Studies in Honour of Stig Johansson*, a cura di Karin Aijmer and Bengt Altenberg, John Benjamins Publishing Company, 2013.
- ENGHELS-DEFRANCQ-JANSEGGERS 2020 = RENATA ENGHELS, BART DEFRANCQ, MARLIES JANSEGGERS, *Reflections on the Use of Data and Methods in Contrastive Linguistics*, in *New Approaches to Contrastive Linguistics: Empirical and Methodological Challenges*, a cura di Renata Enghels, Bart Defrancq e Marlies Jansegers, 2020, pp. 1-20.
- FILLMORE 1970 = CHARLES J. FILLMORE, *The Grammar of Hitting and Breaking*, in *Readings in English Transformational Grammar*, a cura di Roderick Jacobs e Peter Rosenbaum, Ginn and company, 1970, pp. 9-29.
- FUSCO 2006 = FABIANA FUSCO, *La Traduttologia: Concetti e Termini*, Udine, Forum, 2006.
- GAST 2012 = VOLKER GAST, *Contrastive Linguistics: Theories and Methods*, in *Dictionaries of Linguistics and Communication Science: Linguistic Theory and Methodology*, Berlin, Mouton de Gruyter, 2012.
- GILQUIN 2000 = GAËTANELLE GILQUIN, *The Integrated Contrastive Model: Spicing up Your Data*, in «Languages in Contrast», 3 (1), 2000, pp. 95-123.
- GODDARD-WIERZBICKA 2009 = CLIFF GODDARD, ANNA WIERZBICKA, *Contrastive Semantics of Physical Activity Verbs: 'Cutting' and 'Chopping' in English, Polish, and Japanese*, in «Language Sciences», 31 (1), 2009, pp. 60-96.
- GRANGER-LEROT-PETCH-TYSON 2003 = SYLVIANE GRANGER, JACQUE LEROT, STEPHANIE PETCH-TYSON, *Corpus-Based approaches to Contrastive Linguistics*, Leiden, The Netherlands, Brill, 2003.
- GRIES-JANSEGGERS-MIGLIO 2020 = STEPHAN THOMAS GRIES, TH MARLIES JANSEGGERS, VIOLA G. MIGLIO, *Quantitative methods for corpus-based contrastive linguistics*, in R. Enghels, B. Defrancq, & M. Jansegers (Eds.), *New approaches to contrastive linguistics: empirical and methodological challenges*, vol. 336, 2020, pp. 53-84.
- GRIES 2021 = STEFAN THOMAS GRIES, *Statistics for Linguistics with R: A Practical Introduction*, Berlin, Boston, De Gruyter Mouton, 2021.
- HASPELMATH 2010 = MARTIN HASPELMATH, *Comparative Concepts and Descriptive Categories in Crosslinguistic Studies*, in «Language», 86 (3), 2010, pp. 663-687.

- HASSELGÅRD 2020 = HILDE HASSELGÅRD, *Corpus-Based Contrastive Studies: Beginnings, Developments and Directions*, In *The Complementary Contribution of Comparable and Parallel Corpora to Crosslinguistic Studies*, a cura di Sylviane Granger and Marie-Aude Lefer, John Benjamins Publishing Company, 2020.
- JOHANSSON 2007 = STIG JOHANSSON, *Seeing through Multilingual Corpora*, in «Corpus Linguistics 25 Years on», a cura di Roberta Facchinetti, Leiden, The Netherlands, Brill, 2007, pp. 51-71.
- KOEHN 2005 = KOEHN PHILIPP, *Europarl: A parallel corpus for statistical machine translation*, in *Proceedings of Machine Translation Summit X: Papers*, 2005, pp. 79-86.
- KÖNIG 2012 = EKKEHARD KÖNIG, *Contrastive Linguistics and Language Comparison*, in «Languages in Contrast», 12 (1), 2012, pp. 3-26.
- KOPECKA-NARASIMHAN 2012 = ANETTA KOPECKA, BHUVANA NARASIMHAN, *Events of Putting and Taking: A Crosslinguistic Perspective*, Amsterdam, John Benjamins, 2012.
- KRZESZOWSKI 1990 = TOMASZ P. KRZESZOWSKI, *Contrasting Languages: The Scope of Contrastive Linguistics*, De Gruyter, 1990.
- LEVSHINA 2015 = NATALIA LEVSHINA, *How to Do Linguistics with R*, Amsterdam/Philadelphia, John Benjamins Publishing Company, 2015.
- MAJID-JORDAN-DUNN 2015 = ASIFA MAJID, FIONA JORDAN, MICHAEL DUNN, *Semantic Systems in Closely Related Languages*, in «Language Sciences», 49, 2015, pp. 1-18.
- MAJID-BOSTER-BOWERMAN 2008 = ASIFA MAJID, JAMES S. BOSTER, MELISSA BOWERMAN, *The Cross-Linguistic Categorization of Everyday Events: A Study of Cutting and Breaking*, in «Cognition», 109 (2), 2008, pp. 235-250.
- MAJID ET AL. 2007 = ASIFA MAJID, MARIANNE GULLBERG, MIRIAM VAN STADEN, MELISSA BOWERMAN, *How Similar Are Semantic Categories in Closely Related Languages? A Comparison of Cutting and Breaking in Four Germanic Languages*, in «Cognitive Linguistics», 18 (2), 2007, pp. 179-194.
- MALT 2010 = BARBARA C. MALT, *Naming Artifacts: Patterns and Processes*, in «The Psychology of Learning and Motivation», 52, 2010, pp. 1-38.
- MAURANEN 1999 = ANNA MAURANEN, *Will 'translationese' Ruin a Contrastive Study?*, in «Languages in Contrast», 2 (2), 1999, pp. 161-185.
- MONEGLIA 1993 = MASSIMO MONEGLIA, *La Sottodeterminazione Del Significato Lessicale e l'equiestensionalità Locale Nel Paradigma Di «aprire»*, in «Studi di Grammatica Italiana», 15, 1993, pp. 297-324.

## Caterina Cacioli

- MONEGLIA 2000 = MASSIMO MONEGLIA, *Cambiamenti Semantici Palesi e Nasosti Nel Lessico Verbale Italiano: Le Transizioni Di Possesso*, in «Lingua e Stile», XXXV, 4, 2000, pp. 629-664.
- MONEGLIA ET AL. 2012 = MASSIMO MONEGLIA, MONICA MONACHINI, OMAR CALABRESE, ALESSANDRO PANUNZI, FRANCESCA FRONTINI, GLORIA GAGLIARDI, IRENE RUSSO, *The IMAGACT Cross-Linguistic Ontology of Action. A New Infrastructure for Natural Language Disambiguation*, in *Proceedings of the Eight International Conference on Language Resources and Evaluation*, Paris, European Language Resources Association (ELRA), 2012, pp. 948-955.
- MONEGLIA ET AL. 2014 = MASSIMO MONEGLIA, SUSAN W. BROWN, FRONTINI FRANCESCA, GLORIA GAGLIARDI, KHAN FAHAD, MONICA MONACHINI, ALESSANDRO PANUNZI, *The IMAGACT Visual Ontology. An Extendable Multilingual Infrastructure for the Representation of Lexical Encoding of Action*, in *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014*, 3425-32, European Language Resources Association (ELRA), 2014.
- NORDRUM-EBELING-HASSELGÅRD 2016 = LENE NORDRUM, SIGNE OKSEFJELL EBELING, HILDE HASSELGÅRD, *Introduction-Languages in Contrast 20 Years On*, in «NJES Nordic Journal of English Studies», 15 (3), 2016, pp. 1-6.
- PYE 1996 = CLIFTON PYE, *K'iche' Maya Verbs of Breaking and Cutting*, in «Kansas Working Papers in Linguistics», 21, 1996, pp. 87-98.
- VIBERG 2013 = ÅKE VIBERG, *Posture Verbs: A Multilingual Contrastive Study*, in «Languages in Contrast», 13 (2), 2013, pp. 139-169.
- VIBERG 2020 = ÅKE VIBERG, *Contrasting Semantic Fields across Languages*, in *New Approaches to Contrastive Linguistics: Empirical and Methodological Challenges*, De Gruyter, 2020, pp. 265-312.
- VINOGRADOV 2023 = IGOR VINOGRADOV, *Pioneers of Contrastive Linguistics. Dominican Missionaries in Highland Guatemala*, in «Languages in Contrast»; 23 (1), 2023, pp. 34-59.

## Sitografia

Imagact = [www.imagact.it](http://www.imagact.it)

Sketchengine = [www.sketchengine.eu](http://www.sketchengine.eu)

## Appendice

1. Rompere una bottiglia sbattendola su un muretto
2. Forzare l'apertura di una cassaforte
3. Colpire una scatola fino a distruggere interamente la sua forma originaria
4. Spezzare una corda tirando le estremità
5. Creare una frattura in uno specchio colpendo con uno strumento
6. Rompere un braccio
7. Schiacciare un uovo con una mano
8. Rompere una sedia separando schienale dalla base
9. Crollo di un palazzo
10. Strappare un foglio in due
11. Strappare un foglio in più parti
12. Dividere una mattonella in tre pezzi facendo pressione con un cacciavite
13. Spezzare un bastoncino in più parti
14. Distruggere i mobili di una stanza
15. Distruggere una macchina facendola scontrare con un muro
16. Staccare, rompendo, una ruota da una macchinina giocattolo
17. Staccare un ramo da una pianta
18. Sgusciare delle noccioline
19. Rompere un uovo
20. Spezzare una matita in due
21. Rompere delle noci
22. Dare una spallata a una porta chiusa a chiave e sondare la porta
23. Colpire una scatola di legno con un martello, creando un foro
24. Schiacciare una lattina con il piede
25. Rompere un ventilatore inserendo uno strumento nel meccanismo
26. Colpire un monitor

## Caterina Cacioli

**Riassunto** La ricerca in semantica contrastiva si trova ad affrontare sfide metodologiche sulle risorse usate, la comparabilità dei dati interlinguistici e i criteri dei termini da contrastare. In questo lavoro si tratteranno le questioni menzionate per proporre l'uso combinato di una risorsa linguistica multimodale (Imagact) e di *corpora* per la ricerca in semantica contrastiva dei verbi d'azione. Nel caso di studio qui presentato si mostreranno i vantaggi della combinazione di risorse e di dati linguistici, il contributo della componente visuale della risorsa Imagact risulti nella disambiguazione dei livelli concettuali e linguistici del significato, e l'informazione suppletiva che forniscono i *corpora*.

**Abstract** Research in contrastive semantics have found methodological challenges along its way, related to the resources used, the comparability of inter-linguistic data and the criteria chosen for comparison. This article will tackle these issues and propose the combination of a multimodal, multilingual resource (Imagact) with *corpora* in the contrastive analysis of action verbs. In the case study presented here, it will be shown which are the advantages of combining resources and linguistic data, how the visual component of the Imagact resource helps disambiguating the conceptual and linguistic levels of meaning-making, and the complementing information provided by *corpora*.