



UNIVERSITÀ DI PISA

Dipartimento di Filologia, Letteratura e Linguistica

Corso di Laurea Magistrale in Linguistica

Tesi di Laurea

Il fenomeno della produttività sintattica nella selezione
argomentale: un'indagine sperimentale.

CANDIDATA: Beatrice Bruni

RELATORE:

Chiar.mo Prof. Alessandro Lenci

CONTRORELATORE:

Chiar.mo Prof. Roberto Peroni

ANNO ACCADEMICO 2015/2016

Sommario.

Introduzione.	1
1. FONDAMENTI TEORICI: LE STRUTTURE ARGOMENTALI.	5
1.1 Approccio proiezionista.	7
1.1.1 I limiti degli approcci proiezionisti.	11
1.2 Le grammatiche costruzioniste e i modelli usage-based del linguaggio.	13
1.3 I modelli usage-based del linguaggio.	17
1.3.1 Studi a sostegno dei modelli usage-based del linguaggio.	21
1.4 Sintesi.	25
2. PRODUTTIVITÀ SINTATTICA: LO STATO DELL'ARTE.	26
2.1 Le nozioni di produttività, creatività e analogia fra generativismo e modelli usage-based del linguaggio.	26
2.2 Produttività: dalla morfologia alla sintassi.	32
2.2.1 Definizioni e misure di produttività morfologica	32
2.2.2 Dalla morfologia alla sintassi	35
2.2.3 Definizioni di produttività sintattica.	39
2.3 Fattori di produttività sintattica nelle costruzioni a struttura argomentale.	43
2.3.1 Produttività, semantica lessicale e conoscenza del mondo.	43
2.3.2 Misure di produttività sintattica.	47
2.3.3 Studi precedenti.	52
2.3.4 Una rappresentazione cognitiva della produttività sintattica: la legge di Hebb.	60
2.4 Sintesi.	62
3. OBIETTIVI, MATERIALI E CRITERI DI ANALISI.	64
3.1 Obiettivi della ricerca.	64

3.2	La scelta dei verbi.	67
3.2.1	Il problema della sinonimia in linguistica.	67
3.2.2	Criteri di selezione dei verbi.	69
3.2.3	Strumenti per la selezione dei verbi.	70
3.3	Strumenti di indagine e il corpus di riferimento.	71
3.3.1	Lexit.	71
3.3.2	Il corpus La Repubblica.	72
3.4	Estrazione dei dati da Lexit.	73
3.5	Analisi dei dati estratti.	75
3.5.1	Frequenza Token, frequenza type, type-token ratio.	77
3.5.2	Differenze nella selezione argomentale.	85
3.6	Sintesi	99
4.	REALIZZAZIONE DEL TEST, ANALISI DEI RISULTATI, PROSPETTIVE FUTURE.	101
4.1	La realizzazione del test.	101
4.1.1	Procedura.	101
4.1.2	Il metodo del crowdsourcing.	105
4.1.3	I partecipanti.	107
4.2	I risultati.	107
4.3	Discussione.	132
4.4	Prospettive future.	136
	Conclusioni.	138
	Bibliografia.	141
	Sitografia.	148
	Appendice.	149

Introduzione.

Il presente studio si propone l'obiettivo di analizzare, attraverso un'indagine di carattere sperimentale, il fenomeno della produttività sintattica nella selezione argomentale da parte di coppie di verbi quasi sinonimi italiani.

Il fenomeno della produttività in linguistica, per la sua natura complessa e sfaccettata, ha da sempre attirato l'attenzione di numerosi studiosi che, attraverso prospettive diverse, spesso opposte tra loro, hanno cercato di indagare in che modo la lingua cambia e si evolve per andare in contro alle esigenze comunicative dei parlanti. Della produttività si è parlato in diacronia, prendendo il tempo come fattore responsabile del cambiamento ma anche attraverso una prospettiva sincronica, analizzando in che modo le costruzioni vadano in contro ad un fenomeno di generalizzazione che le porta ad essere estese a nuovi usi.

Gli studi sulla produttività nascono in ambito morfologico, e hanno riguardato i fenomeni di formazione e derivazione delle parole per estendersi poi, in un secondo momento, all'ambito sintattico. Numerose sono le teorie che spesso in contraddizione fra loro, hanno cercato di darne una definizione e di stabilire dei criteri d'indagine che permettessero di indagare più a fondo il fenomeno nei suoi diversi aspetti.

Per quanto riguarda la presente indagine, l'orizzonte teorico di riferimento è costituito dalle grammatiche costruzioniste di stampo usage-based, una famiglia di teorie che postulano alla base di qualsiasi indagine linguistica la nozione di costruzione, come sinolo di forma e significato a vari livelli di astrazione. Secondo tali modelli, la comprensione di qualsiasi oggetto linguistico non può esulare dall'osservazione degli usi effettivi che i parlanti fanno della lingua, la cui rappresentazione cognitiva, nella mente del parlante, è dinamica e pronta al cambiamento.

Secondo tale prospettiva, ogni costruzione è caratterizzata da parti lessicalmente specificate e da slot, posizioni aperte che possono accogliere al loro interno vario materiale lessicale nel rispetto dei vincoli semantici imposti dalla costruzione stessa.

Secondo tale prospettiva, il concetto di produttività combacia con quello di *estendibilità* (Barðdal, 2008) e si riferisce alla capacità di un pattern di accettare nuovi items all'interno dei propri slot. È proprio la schematicità degli slot di una costruzione che permette ai procedimenti analogici di intervenire: più i suoi componenti sono flessibili più essa sarà produttiva.

La possibilità che una costruzione venga estesa a nuovi elementi lessicali, dipende inoltre, da come le istanze testimoniate sono distribuite nella lingua e da come, il potenziale nuovo item è collegato a queste ultime, oltre che all'esistenza o meno di un'alternativa già presente e che statisticamente sia in grado di vincolare tale estensione. La produttività è dunque analizzata come un fenomeno scalare e graduato, pertanto, l'obiettivo di questi modelli, è, più che darne una definizione univoca, quello di quantificarne la portata.

La nostra indagine si inserisce all'interno di questo scenario, e si aggancia, in particolare, a uno studio condotto da Amir Zeldes nel 2012 dal titolo "Productive argument selection: Is lexical semantic enough?", nel quale lo studioso si propone di analizzare, a livello statistico, i diversi pattern di produttività di costruzioni sinonimiche, nel fenomeno della selezione argomentale.

Come Zeldes abbiamo deciso di lavorare su coppie di verbi quasi-sinonimi per indagare il grado di generalizzazione degli stessi e verificare se, e in quale misura, uno dei due membri presentasse una maggiore predisposizione rispetto all'altro ad accettare nuovo materiale lessicale all'interno dello slot argomentale.

Entrambe si caratterizzano per la natura corpus-based: sono sempre più numerosi gli studi di linguistica, che fanno uso di corpora di riferimento di lingua scritta e parlata per analizzare determinati fenomeni linguistici.

Se però, Zeldes si ferma ad un'analisi di tipo statistico-computazionale dei dati estratti, il nostro obiettivo è stato quello di utilizzare i dati per creare un task di natura psicolinguistica. I dati estratti sono dunque serviti per suggerire determinate ipotesi di partenza che abbiamo voluto verificare sperimentalmente, sui parlanti.

Il lavoro è organizzato nel modo seguente.

Nel primo capitolo, abbiamo delineato gli orizzonti teorici di riferimento, parlando in particolare della Construction Grammar e delle teorie usage-based del linguaggio, analizzate in contrapposizione alle teorie proiezioniste di stampo generativista. Abbiamo delineato i capisaldi di tali approcci e descritto brevemente alcuni degli studi più importanti che sono stati compiuti nell'ambito della grammatica costruzionista di stampo usage-based, per offrire evidenze sperimentali a supporto delle tesi proposte.

Nel secondo capitolo, abbiamo stabilito l'oggetto stesso dello studio, la produttività sintattica, contestualizzandolo all'interno dell'orizzonte teorico di riferimento. Abbiamo tracciato dunque lo stato dell'arte degli studi condotti riguardo al fenomeno in termini

usage-based, evidenziando quali sono i punti condivisi e quelli di distacco. In particolare ci siamo soffermati sui lavori di Goldberg (2006), Barðdal (2008), Zeldes (2012), Zeschel (2012) e Perek (2015).

Nel terzo capitolo,

Siamo poi passati a delineare i vari passaggi della presente indagine: abbiamo deciso di analizzare coppie di verbi italiani quasi-sinonimi per verificare se e in quale misura presentassero tendenze divergenti nella selezione argomentale e nella generalizzazione a nuovi argomenti.

Nel terzo capitolo, è descritta l'estrazione e l'analisi dei dati. Pattern V+N, per ciascuna coppia di sinonimi, sono stati estratti da Lexit, un framework computazionale che permette l'esplorazione dei profili distribuzionali di nomi, verbi e aggettivi italiani e che fa capo al corpus La Repubblica. Insieme alle coppie V+N sono state raccolte informazioni circa la frequenza di ricorrenza di ciascuna di esse nel corpus. Per ogni pattern è stata calcolata la frequenza token, la frequenza type, il numero di hapax legomena e la type/token ratio come indice di ricchezza e varietà lessicale.

Attraverso questi dati ci è stato permesso di fare ipotesi circa l'uso che dei verbi è fatto nella lingua e riguardo alla loro predisposizione ad essere estesi a nuovo materiale lessicale.

Le varie di coppie di sinonimi sono state poi confrontate per quanto riguarda la scelta degli argomenti selezionati, per verificare se all'interno della coppia si verificassero preferenze di selezione idiosincratiche per ciascun membro.

Una volta delineati i casi più interessanti, siamo passati alla realizzazione dell'esperimento per il quale sono state selezionate dieci coppie di verbi sinonimici.

Nel quarto capitolo, sono descritte la realizzazione del design sperimentale, la creazione degli stimoli e la metodologia utilizzata per implementare l'esperimento.

Per condurre il task, nella presentazione degli stimoli e nella raccolta dei risultati, è stato utilizzato *crowdfower*, una piattaforma crowdsourcing creata nel 2007 in California, che ci ha permesso di raccogliere in breve tempo le risposte necessarie a condurre la nostra indagine.

Per ogni coppia di sinonimi sono stati selezionati determinati nomi sulla base di tre categorie di frequenza: argomenti che ricorrevano con frequenza elevata con entrambi i verbi, argomenti che ricorrevano con frequenza elevata con un verbo ma bassa con l'altro e viceversa, e argomenti che ricorressero solo con uno dei due verbi e non con

l'altro e viceversa. Sulla base di queste caratteristiche sono stati creati gli stimoli, rappresentati dal verbo all'infinito seguito dall'articolo e dal nome (es. *incorrere in un incidente*). Questi ultimi sono stati presentati ai soggetti ai quali è stato chiesto, per ogni espressione così formata, di fornire un giudizio di accettabilità sulla base di una scala likert da 1 a 7.

I risultati raccolti sono stati sottoposti ad analisi statistiche, il cui esito è discusso nella parte finale del capitolo.

Il capitolo si conclude con alcune riflessioni riguardo a possibili prospettive future e consigli per un eventuale follow up.

Per concludere, il presente studio si propone di chiarire le dinamiche di generalizzazione dei verbi nel processo di selezione argomentale, e di evidenziare come la semantica lessicale non sia l'unico fattore in grado di fornire una spiegazione esaustiva del fenomeno. Numerosi sono gli elementi che entrano in gioco e influenzano la propensione di una costruzione ad essere estesa a nuovo materiale lessicale, ed essi hanno a che fare in gran parte, con l'esperienza che il parlante fa ogni giorno della lingua.

1. Fondamenti teorici: le strutture argomentali.

Ogni lingua codifica diversamente il modo in cui i parlanti parlano degli eventi e di coloro che ne prendono parte. Tale funzione è in larga misura svolta dal verbo il quale, a causa della sua natura inerentemente relazionale, spesso compare accompagnato da altre parole, gli argomenti.

Di tutti gli elementi che compaiono all'interno della frase, il verbo, è quello che porta il maggiore contributo informativo sia dal punto di vista sintattico che semantico.

Negli anni sessanta, Tesnière (1959), padre della Grammatica Valenziale, riprendendo un concetto della chimica, aveva parlato di valenza verbale: un verbo, esattamente come un atomo, si lega ad altri elementi (gli attanti) e il numero di tali elementi (valenza) è determinato dal verbo stesso. Esistono quindi verbi zerovalenti (avalenti), monovalenti, bivalenti ecc., a seconda del numero di argomenti che un verbo richiede per essere "saturato". I complementi non richiesti dal verbo sono invece detti "circostanziali", e indicano le circostanze in cui si svolge l'azione.

Tesnière, in termini metaforici, aveva paragonato il verbo, nucleo centrale della frase, a un dramma teatrale: come il dramma, il verbo mette in scena un'azione a cui partecipano degli attori che nella scena svolgono un ruolo ben definito.

Nella sua analisi, egli tuttavia, non prende in considerazione la natura semantica e la realizzazione sintattica degli argomenti che considera idiosincratica e imprevedibile.

L'importante intuizione di Tesnière viene successivamente arricchita da ulteriori riflessioni teoriche e la nozione di valenza viene sostituita progressivamente da quella di "struttura argomentale"; la semplice indicazione del numero degli attanti viene affinata con nozioni relative alla realizzazione sintattica e al ruolo semantico, detto anche tematico, degli argomenti.

I parlanti sanno che ogni argomento, affinché un'espressione abbia un determinato significato, deve apparire in una precisa posizione all'interno della frase e svolgere un determinato ruolo. Il verbo *uccidere* ad esempio, è solitamente accompagnato da due argomenti, uno con ruolo di *agente*, realizzato sintatticamente come soggetto, l'altro con il ruolo di *paziente*, realizzato sintatticamente come complemento oggetto diretto. Tali argomenti inoltre devono comparire in specifiche posizioni all'interno della frase:

secondo l'ordine delle parole canonico, in italiano, chi compie l'azione occupa la posizione preverbale mentre la vittima quella post-verbale.

1) Bruto uccise Cesare

Tale conoscenza rappresenta la *struttura argomentale* del verbo uccidere.

L'esatta natura di tale conoscenza, e come essa debba essere rappresentata nella grammatica, è oggetto di dibattito nelle varie teorie linguistiche.

È bene sottolineare che i termini *realizzazione argomentale* e *struttura argomentale*, non si riferiscono soltanto allo studio delle espressioni sintattiche degli argomenti di un verbo ma possono essere applicati ad ogni parola in grado di governare altri elementi all'interno di una frase. Il fatto che la letteratura si sia concentrata soprattutto sullo studio dei verbi è dovuto al fatto che, come notano Levin e Rappaport Hovav (2005:33), questi ultimi costituiscono “the prototypical predicators, that is, argument-taking words”.

La struttura argomentale ha due facce, una semantica, che definisce i partecipanti all'evento espresso dal verbo e una sintattica, che stabilisce la rappresentazione gerarchica degli argomenti richiesti dal predicato e che determina come essi debbano essere rappresentati in sintassi.

Dagli anni ottanta molte teorie grammaticali si sono basate sull'assunto che la realizzazione sintattica degli argomenti, la loro categoria e la loro funzione grammaticale, fosse predicibile a partire dal significato dei verbi. Queste teorie considerano la struttura sintattica come proiezione delle proprietà lessicali del predicato, espresse in termini di ruoli tematici, di proto-ruoli e in termini di scomposizione lessicale del predicato.

Esiste tuttavia una serie di fenomeni linguistici che ha portato a mettere in dubbio il fatto che la sintassi e la semantica della struttura argomentale fosse proiettata esclusivamente dal verbo. In particolare è stato notato come i verbi possano ricorrere in più di una struttura argomentale accompagnati da una sistematica variazione di significato. L'elasticità dei verbi permette loro di assumere significati diversi a seconda degli argomenti con cui appaiono.

Studi recenti, hanno sottolineato che la conoscenza riguardo alle strutture argomentali non possa essere unicamente legata al verbo ma coinvolga una serie di principi generali

indispensabili per spiegare determinati fenomeni linguistici come gli errori di (iper-)generalizzazione compiuti in età prescolare dai bambini, o l'uso creativo delle strutture argomentali ad opera dei parlanti adulti per soddisfare determinate esigenze comunicative.

Tali fenomeni sono difficili da spiegare partendo dall'idea che tutta la conoscenza riguardo alla struttura argomentale sia racchiusa unicamente nel verbo.

Questo tipo di osservazioni hanno portato alcuni studiosi a ritenere che parte dell'informazione semantica fosse associata direttamente alla struttura argomentale astratta in cui compare il verbo, riconoscendo tale struttura argomentale come unità linguistica a sé stante e chiamata costruzione, e a sottolineare l'importanza svolta dall'esperienza linguistica del parlante nell'estendere l'uso di determinati verbi a nuovi utilizzi.

In questo capitolo mi concentrerò su alcune delle problematiche teoriche affrontate dai recenti approcci riguardo alla struttura argomentale i quali si sono chiesti se il comportamento di un verbo rispetto ai suoi argomenti fosse del tutto idiosincratico, o se invece fosse possibile estrapolare delle regole generali, ossia delle componenti nel significato del verbo che concorressero a determinarne il comportamento sintattico, costituendo così delle classi lessicali omogenee; se e fino a che punto la struttura argomentale possa essere considerata esclusivamente una proprietà lessicale del verbo o se piuttosto non sia sensibile anche al contesto, alla concettualizzazione dell'evento o addirittura se non sia la struttura stessa ad essere portatrice di significato. In questo capitolo esporremo le linee generali dei due principali modelli teorici che hanno affrontato l'argomento: l'approccio proiezionista e quello costruzionista; concluderemo facendo riferimento ai moderni modelli linguistici basati sull'uso, i quali sottolineano l'importanza che riveste l'esperienza linguistica che il parlante ha con la lingua stessa nel creare le rappresentazioni cognitive delle varie strutture linguistiche.

1.1 Approccio proiezionista.

L'idea alla base degli approcci proiezionisti, a partire da Chomsky (1965) è che le strutture argomentali siano proiezioni dell'informazione semantico-lessicale fornita dal verbo la quale, attraverso regole di collegamento universali realizza gli argomenti in

posizioni sintattiche precise.

Tale assunto parte dall'osservazione che i verbi giocano un ruolo di primaria importanza nella conoscenza linguistica dei parlanti e suggerisce, sulla base di studi sperimentali, che siano essi, il principale fattore predittivo del significato della frase.

Healy e Miller (1970) hanno indagato il contributo apportato rispettivamente dal verbo e dal soggetto nel determinare il significato complessivo di una frase. Entrambi gli elementi sembrano essere i candidati migliori a cui attribuire tale valore predittivo: il primo perché veicola l'informazione circa "chi fa cosa a chi", il secondo perché spesso svolge la funzione di *topic* della frase. Gli studiosi hanno realizzato un esperimento i cui stimoli sono stati creati incrociando cinque argomenti soggetto (*the salesman, the writer, the critic, the student, the publisher*) a cinque verbi (*sold, wrote, criticized, studied, published*) e un paziente (*the book*). Ai soggetti è stato chiesto di ordinare le frasi in gruppi sulla base della loro similarità semantica. I risultati dell'esperimento mostrano che i partecipanti hanno una maggiore predisposizione ad ordinare insieme frasi caratterizzate dallo stesso verbo piuttosto che quelle contenenti lo stesso soggetto; i due studiosi ne concludono che sia il verbo l'elemento predittivo principale del significato di una frase (cfr. Cao, 2014 per una visione analoga del ruolo svolto dal verbo nella comprensione di frasi inglesi ad opera di parlanti cinesi).

Altri studi psicolinguistici sembrano mostrare che le preferenze statistiche di un verbo per una determinata struttura argomentale svolgano un ruolo importante nella comprensione on-line di frasi ambigue (cfr. Garnsey, Pearlmutter, Myers e Lotocky 1997, Trueswell, Tanenhaus e Kello 1993).

Grazie anche al contributo di questi studi, si sono andate consolidando le cosiddette teorie proiezioniste della realizzazione argomentale, secondo le quali, la conoscenza della struttura argomentale è legata al verbo, il quale stabilisce il numero e il tipo di argomenti con cui apparire.

Alla base di tali modelli risiede l'idea che il comportamento sintattico dei verbi sia determinato da "aspetti grammaticalmente rilevanti" del suo significato, cioè che solo determinati aspetti del significato siano rilevanti a livello della realizzazione argomentale e che l'occorrenza di certi verbi, in determinate cornici sintattiche e non in altre, possa essere spiegato con la presenza/assenza di questi tratti semantici distintivi.

A tale riguardo Pinker (1989) afferma che:

Perhaps there is a set of semantic elements and relations that is much smaller than the set of cognitively available and culturally salient distinctions, and verb meanings are organized around them. (Pinker, 1989: 166)

Tali componenti sono isolate esaminando il denominatore semantico comune a quei verbi che esibiscono le stesse opzioni di realizzazione argomentale.

Uno dei contributi principali a tale riguardo, risale al lavoro di Levin (1993) la quale ha raggruppato 3000 verbi inglesi in classi semantiche coerenti sulla base del loro comportamento sintattico e delle loro proprietà semantiche.

Secondo Levin (1993), se è vero che non esiste una relazione deterministica tra significato e struttura argomentale, d'altra parte è vero anche che vi sono delle componenti nel significato lessicale che sono alla base di un certo comportamento sintattico, comune ai verbi che le possiedono.

Gli elementi del significato che servono a distinguere le varie classi verbali costituiscono l'interfaccia fra semantica e sintassi, sono universali e devono essere in numero relativamente ridotto.

Tali teorie riducono dunque i significati dei verbi a strutture semantiche essenziali composte da un insieme limitato di primitivi semantici e postulano l'esistenza di regole di collegamento generali (*linking rules*) in grado di predire in modo deterministico la realizzazione sintattica degli argomenti di un verbo sulla base della sua struttura semantica.

Levin e Rappaport (2005), hanno proposto un modello per cui i predicati sono scomposti in strutture eventive primitive. Secondo tale modello, l'informazione lessicale sarebbe articolata su due livelli: una struttura eventiva, che si trova ad un livello semantico-lessicale e una struttura argomentale, collocata ad un livello sintattico-lessicale. La prima fornisce una scomposizione del significato lessicale a livello di struttura semantica, la seconda invece è responsabile del numero di argomenti selezionati dal predicato e della gerarchia che può essere stabilita fra di essi. In questo modo, la realizzazione sintattica degli argomenti ha luogo in base ad una serie di regole che mappano la rappresentazione semantica in quella sintattica. La scomposizione semantica dei verbi causativi, come *rompere*, può essere rappresentata nel modo seguente:

[[X ACT] CAUSE [Y BECOME <BROKEN>]

Il significato verbale è scomposto in due fattori: uno di tipo strutturale ed uno di tipo idiosincratico. La prima componente (*event structure* o *event schema*) è grammaticalmente rilevante nella realizzazione argomentale ed è comune a tutti i verbi che fanno parte della medesima classe verbale; la seconda distingue il verbo in questione dagli altri verbi appartenenti alla stessa classe, ma non è rivelante nel determinare il comportamento grammaticale del verbo (*root*). Ogni *root* ha una caratterizzazione di tipo ontologico (*state, result state, thing, stuff, surface/container, manner, instrument*) la quale determina la sua associazione con la *event structure*.

L'integrazione fra la *root* e gli schemi di evento è regolata dalle *Canonical Realization Rules* (Rappaport Hovav e Levin, 1998) e può avvenire in due modi, riempiendo una posizione argomentale associata con il predicato primitivo (2a) e (2b) oppure fungendo da modificatore del predicato stesso (3)

- 2) a. *dry*: [[X ACT] CAUSE [Y BECOME < DRY >]]
 b. *bottle*: [[X ACT] CAUSE [Y BECOME IN < BOTTLE >]]
- 3) *jog*: [X ACT < JOG >]

I verbi che appartengono alla medesima classe semantica presentano una scomposizione semantica caratterizzata dalle medesime sotto-strutture, nella quale le *root* dello stesso tipo ontologico riempiono le medesime posizioni, e condividono anche le proprietà sintattiche rilevanti nella determinazione della realizzazione argomentale.

Molti dei modelli di scomposizione semantica, hanno posto alla base della realizzazione argomentale, varie versioni del *principio di realizzazione argomentale* per cui:

- a. There must be an argument XP in the syntax for each structure participant in the event structure.
- b. Each argument XP in the syntax must be associated with an identified sub-event in the event structure.

A cui Levin e Rappaport (1998) aggiungono un'ulteriore condizione:

Subevent Identification Condition: each sub-event in the event structure must be identified by a lexical head (e.g., a V, an A, or a P) in the syntax. (Rappaport Hovav e Levin, 1998: 112)

Entrambe partono dal presupposto che gli elementi appartenenti alla rappresentazione semantica debbano ricevere una codifica a livello sintattico e che la rappresentazione sintattica, a sua volta, sia pienamente interpretabile a livello semantico.

Sebbene i vari approcci proiezionisti differiscano per quanto riguarda il tipo di primitivi semantici impiegati, la forma assegnata alla struttura semantica e i principi alla base del mapping fra semantica e sintassi, l'elemento che comunque rimane comune a tutti i modelli è l'idea che la struttura argomentale sia interamente determinata dalle proprietà del verbo e che nel processo di mappatura la sintassi preservi determinate distinzioni semantiche.

1.1.1 I limiti degli approcci proiezionisti.

Studi recenti hanno fatto emergere alcuni limiti degli approcci proiezionisti. Essi riguardano un aspetto interessante della realizzazione argomentale, ovvero il fatto che il significato del verbo muta a seconda del frame sintattico in cui esso è inserito.

Spesso capita, inoltre, che i parlanti utilizzino i verbi in modo non convenzionale, all'interno di contesti sintattici atipici.

4) She sneezed the foam off the cappuccino. (Goldberg 2006)

Se, in linea con gli approcci proiezionisti, si ritiene che la realizzazione argomentale, possa essere predetta in modo deterministico dalla rappresentazione semantico-lessicale del verbo, come si spiega il fatto che lo stesso verbo possa avere realizzazioni sintattiche differenti?

Gli approcci proiezionisti hanno spiegato questo fenomeno postulando l'esistenza di regole lessicali che prendono in input un verbo caratterizzato da una determinata valenza e un significato basico e restituiscono in output un verbo con una valenza diversa ed un significato esteso. Con lo scopo di preservare un rapporto uno-a-uno fra struttura semantica e struttura argomentale, tali modelli fanno ricorso al concetto di

polisemia verbale, ritendendo che ogni volta che il verbo compare in una struttura sintattica diversa, esso sia rappresentato anche da una differente struttura semantica.

Lexical rule-based approach assume a lexical rule that take an intransitive (version of a) verb as input and licenses a special lexical item that selects for an additional object and a secondary predicate. (Muller, 2006:23)

Nel caso specifico di *sneeze*, viene postulata l'esistenza di una regola che trasforma il verbo intransitivo in uno che richieda un oggetto diretto e uno obliquo, col significato "x causes y to move (to) z by sneezing" e proietti l'agente come soggetto, il tema come oggetto e il sintagma preposizionale come obliquo.

5) *sneeze* <agent> → *sneeze* <agent, theme, path>

È stato notato, tuttavia, come la pervasività del fenomeno porterebbe inevitabilmente a presupporre l'esistenza di una serie di sensi creati ad hoc per giustificare l'occorrenza dei verbi nelle varie cornici sintattiche.

Esistono inoltre casi in cui differenze nella realizzazione argomentale non sono correlate a differenze nel significato. Gli studiosi si sono spesso domandati se determinate alternanze argomentali fossero associate a differenze semantiche o se non fossero piuttosto il riflesso di una diversa modalità di "*information packaging*" proprio perché alcune di esse sembrano alternare solo nell'espressione morfosintattica degli argomenti senza una vera e propria variazione di significato.

I problemi che derivano da questi approcci, hanno portato in anni recenti allo sviluppo di teorie linguistiche che rifiutano l'idea che l'informazione riguardo alla realizzazione argomentale sia unicamente veicolata dal verbo ma al contrario, determinati aspetti della forma e del significato delle frasi siano portate dalle costruzioni sintattiche stesse. Questi modelli prendono il nome di approcci costruzionisti.

1.2 Le grammatiche costruzioniste e i modelli usage-based del linguaggio.

Le grammatiche *costruzioniste* affondano le proprie radici nei lavori di stampo cognitivista dei linguisti americani: William Croft, Charles Fillmore, George Lakoff, Ronald W. Langacker.

Con il termine *costruzionista* si intende, da una parte, delineare una famiglia di teorie che pone al centro della propria indagine la nozione di *costruzione*, una corrispondenza di forma e significato, come unità basilare della lingua, dall'altra, si vuole anche evidenziare il fatto che la conoscenza linguistica è *costruita*, dal parlante, sulla base dell'input e di vincoli pragmatici e cognitivi.

Le grammatiche costruzioniste si presentano come teorie grammaticali di tipo olistico volte ad analizzare tutti i fenomeni della lingua, distaccandosi, in questo modo, dalle teorie di stampo generativista che proponevano una visione dicotomica della lingua basata su una netta distinzione fra lessico e sintassi.

La conoscenza linguistica è dunque analizzata come un vasto repertorio di segni, ognuno caratterizzato dallo stesso formato.

All levels of grammatical analysis involve constructions: learned pairings of form with semantic or discourse function, including morphemes or words, idioms, partially lexically filled and fully general phrasal patterns. [...]Both lexical and syntactic constructions are essentially the same type of declarative represented data structure: both pair form with meaning. (Goldberg, 2006:5-7)

La nozione di costruzione, che in un certo senso può essere considerata un'estensione del concetto saussuriano di *segno linguistico*, è preferita a quella di *rule*, propria della Grammatica Generativa, per la quale, secondo una prospettiva di tipo riduzionista, la grammatica, attraverso meccanismi computazionali e compositivi, sarebbe in grado di derivare un numero illimitato di combinazioni a partire da un insieme numericamente stabilito di unità minime.

Since lexical items (memorised bits of language) and grammatical schemas (routinized generalisations over such units) are not seen as objects of a qualitatively different type, syntactic composition is not envisioned as a process of stringing up meaningful words

according to meaningless rules, but rather of “integrating” smaller signs (words) into larger signs (syntactic constructions), i.e. as a process of instantiating the slots of an underspecified construction with concrete lexical material. (Zeschel 2012:8-9)

Si viene così a creare un continuum ai cui estremi troviamo da una parte, gli elementi lessicali pienamente specificati, dall'altra, schemi sintattici astratti mentre, al centro, una vasta gamma di strutture intermedie specificate, a livello lessicale, in misure diverse.

Figura 1 Continuum lessico-sintassi (da Goldberg, 2006:5).

Costruzione	Esempio
Morfema derivativo	<i>pre-, -ing</i>
Parola	<i>anaconda, and</i>
Parola complessa	<i>daredevil, shoo-in</i>
Parola complessa (parzialmente specificata)	[N-s] <i>car-s</i>
Idiom (specificato)	<i>give the Devil his due</i>
Idiom (parzialmente specificato)	<i>jog <someone's> memory</i>
Costruzione covariazionale-condizionale	The Xer the Yer: <i>the more you think about it, the less you understand</i>
Costruzione ditransitiva	Sogg V Ogg1 Ogg2: <i>He baked her a muffin</i>
Passivo	Sogg aux SV _{part. pass} (SP _{by}): <i>The armadillo was hit by a car</i>

Come possiamo notare dalla tabella sopra riportata, le costruzioni contengono uno o più *slot*, ovvero posizioni aperte che possono essere riempite da una vasta gamma di elementi lessicali nel rispetto di vincoli stabiliti dalla costruzione stessa circa la forma specifica che deve assumere il *filler*.

Goldberg nota come, affinché un pattern linguistico sia considerato una costruzione, debba rispondere a particolari prerequisiti: certi aspetti del suo significato o della sua funzione non devono essere strettamente predicibili dalle sue parti componenziali o da altre costruzioni di cui è riconosciuta l'esistenza; anche i pattern pienamente predicibili

possono essere memorizzati come costruzioni purché ricorrano nella lingua con una frequenza elevata.

In questo modo, le costruzioni sono memorizzate, nella mente del parlante, insieme ad informazioni semantiche, pragmatiche, distribuzionali e di *information structure*; comune a molti approcci costruzionisti è l'idea infatti che i parlanti memorizzino anche gli aspetti quantitativi e distribuzionali delle costruzioni e dei contesti in cui esse possono apparire.

Il lessico mentale è così costruito sotto forma di una rete gerarchica in cui costruzioni più specifiche derivano, attraverso legami di eredità (*inheritance links*), da costruzioni più generali.

Lo scopo degli approcci costruzionisti è quello di rendere esplicite tutte le generalizzazioni possibili che un parlante è in grado di formulare. Lo strumento formale che consente la generalizzazione è l'*ereditarietà*, ovvero, la possibilità che una costruzione ha di ereditare alcune caratteristiche da un'altra costruzione più generale.

Rispetto ai modelli lessicalisti che ritenevano che esse fossero proiettate dal verbo, quelli costruzionisti ritengono che le costruzioni siano esse stesse portatrici di significato, indipendentemente dagli *items* specifici che occorrono al loro interno.

In questo modo, vengono superati i due grossi limiti degli approcci proiezionisti: il ricorrere alla nozione di polisemia verbale per spiegare la variazione di senso a cui sono soggetti i verbi quando compaiono all'interno di strutture sintattiche differenti, e l'esistenza di più valenze verbali. Riconoscendo un ruolo primario alla costruzione, è possibile, al contrario, integrare nuovi significati nella semantica della costruzione stessa.

Goldberg sottolinea come nella grammatica non operino esclusivamente meccanismi top-down, per cui sono solo le costruzioni ad imporre il proprio significato; si tratta piuttosto di un'interrelazione fra meccanismi top-down e bottom-up per cui la semantica delle costruzioni e quella dei verbi si fondono in un unico elemento. Questo permette ad una stessa costruzione di poter acquisire numerose sfumature di significato in relazione ai verbi che compaiono al suo interno andando a costituire una rete di definizioni correlate al cui centro si trova un senso centrale, prototipico.

L'occorrenza di un verbo all'interno di una costruzione, è regolata da principi di compatibilità semantica: la semantica del verbo e quella della costruzione si *fondono* per dare vita a costruzioni più complesse

Se da una parte i diversi approcci alla grammatica costruzionista si distinguono in base agli interessi d'indagine, al formalismo utilizzato e a cosa viene identificato con il termine *costruzione*, esistono determinati assunti che sono comuni a tutte.

Tutte si presentano come *non modulari* nel senso che forma e significato sono concepiti come parti di un unico elemento grammaticale, la costruzione, piuttosto che come componenti separate. Le costruzioni si trovano ad ogni livello di analisi linguistica, dai morfemi fino a pattern frasali completamente astratti. Esse adottano un approccio del tipo “what you see is what you get” per cui non vengono postulati livelli sintattici soggiacenti né elementi fonologicamente vuoti.

Sono grammatiche *generative*, richiamando l'accezione originaria del termine utilizzata da Chomsky, in quanto si pongono come obiettivo quello di fornire una rappresentazione formale ed esplicita dei meccanismi espressivi della lingua.

Negli approcci costruzionisti inoltre, la grammatica è vista come un unico modulo (*mono-stratal*), costituito dall'insieme delle costruzioni: non sono necessarie regole di trasformazioni o derivazioni per passare da un modello all'altro. La produzione e l'interpretazione linguistica si basa sulla combinazione di costruzioni differenti, con lo scopo di crearne di nuove. Attraverso l'esposizione all'input linguistico e attraverso capacità cognitive generali, dunque, il parlante estrae strutture linguistiche attraverso le quali interpreta quelle esistenti e ne produce di nuove.

Inoltre, i modelli costruzionisti, si definiscono *full-coverage* poiché si pongono come obiettivo quello di analizzare la totalità dei fenomeni linguistici, da quelli idiosincratici a quelli pienamente regolari, rifiutando la distinzione fra *core grammar* e periferia linguistica tipica della grammatica generativa.

Importante è l'influenza della linguistica cognitiva da cui questi modelli riprendono l'idea che i vincoli del linguaggio derivano da vincoli cognitivi più generali riguardanti la memoria, l'attenzione e da necessità di carattere pragmatico. Molto importante è anche il richiamo alla *social cognition* e alla teoria dell'*embodied cognition* (cfr. Barsalou 2005; Lakoff e Johnson 1999; e altri), per cui, le capacità umane di ragionamento e immaginazione sono strettamente legate all'esperienza senso-motoria con il mondo.

1.3 I modelli usage-based del linguaggio.

“Grammar can be viewed as the cognitive organization of one’s experience with language.” (Bybee, 2013)

Negli ultimi anni è emerso un modello di analisi del linguaggio denominato *usage-based*, per il quale la rappresentazione cognitiva del linguaggio emerge ed è modellata attraverso l’uso della lingua.

Fra i modelli costruzionisti che adottano tale prospettiva ricordiamo la Cognitive Grammar di Langacker (1987, 1991, 2000, 2008), la Cognitive Construction Grammar di Goldberg (1995, 2006), la Radical Construction Grammar di Croft (2001).

Secondo tale approccio la lingua è appresa dai parlanti attraverso l’input linguistico a cui sono continuamente sottoposti senza il bisogno di postulare l’esistenza di una Grammatica Universale innata che vincola e guida gli apprendenti nel processo di acquisizione linguistica e nel suo utilizzo quotidiano. Tutti i processi che contribuiscono a creare determinate costruzioni linguistiche, sono processi generali che si applicano anche ad altri domini cognitivi.

Attraverso tali meccanismi (categorizzazione, associazione cross-modale, *entrenchment*, schematizzazione, analogia, memoria, ecc.), i parlanti codificano le espressioni linguistiche e le categorizzano sulla base della loro forma fonetica, del loro significato, e del contesto in cui appaiono, costruendo in questo modo vere e proprie rappresentazioni cognitive.

Le nuove espressioni sono poi ordinate e combinate per similarità con le rappresentazioni esistenti, facendo emergere le unità linguistiche (sillabe, parole e costruzioni). Ne scaturisce un sistema ben strutturato in cui gli elementi nuovi vengono memorizzati a fianco di unità simili preesistenti e con esse stringono legami di varia natura.

Un ruolo particolarmente importante è svolto dalla memoria. La memoria dell’esperienza linguistica è uguale alla memoria per altri tipi di esperienza: ogni *token* di comportamento linguistico esperito ha un impatto sulla rappresentazione cognitiva e la modifica.

Durante il processo di memorizzazione, vengono immagazzinate informazioni dettagliate relative ai *tokens* linguistici processati e riguardo alla forma e al contesto in cui appaiono.

Secondo Bybee (2010), la memoria è organizzata in unità di processing chiamate *chunks*. Il processo attraverso cui questi blocchi sono generati si riferisce all'abilità cognitiva generale di costruire strutture ricorsivamente, generando l'organizzazione gerarchica della memoria. Le sequenze soggette ad un elevato processi di ripetizione, aderiscono l'una all'altra andando a formare unità più complesse. In questo modo si formano le costruzioni, corrispondenze di forma e significato che vengono impacchettate insieme a livello cognitivo e recuperate come singole unità e non come sequenze di elementi distinti.

Costruzioni ad ogni livello di astrazione sono memorizzate nel lessico mentale il quale contiene simultaneamente sia pattern astratti, sia strutture lessicalmente specificate; le unità di livello più basso, una volta che a partire da esse sono create le generalizzazioni di livello più alto, non vengono abbandonate, ma convivono all'interno del lessico mentale fianco a fianco con esse, in modo gerarchico.

I parlanti formano così una rete complessa di unità che esibiscono vari gradi di radicamento, un inventario fluido e dinamico in grado di adattarsi e modificarsi continuamente in base all'esperienza che il singolo parlante ha con la lingua, senza mai raggiungere una configurazione finale.

Le costruzioni sono memorizzate insieme a informazioni di tipo semantico e distribuzionale, le prime riguardano il frame semantico, la struttura informativa, e nozioni di tipo pragmatico, le seconde, aspetti di tipo quantitativo e distribuzionale grazie alle quali i parlanti sanno in quali contesti linguistici ed extralinguistici, e in quale misura, è opportuno utilizzare una determinata espressione piuttosto che un'altra.

Un ruolo molto importante è attribuito alla frequenza di occorrenza delle forme e delle strutture grammaticali che modifica la loro rappresentazione nella mente dei parlanti.

Uno dei meccanismi fondamentali del linguaggio umano è che le unità e le sequenze di unità linguistiche sono soggette ad un continuo processo di ripetizione che conduce alla loro convenzionalizzazione, a processi di associazione e di automatizzazione.

Ogni token di esperienza ha un impatto sulla memoria: le rappresentazioni cognitive, attraverso il meccanismo della ripetizione, vanno incontro ad un processo di

consolidamento che le rende di più facile accesso. Pertanto, ogni volta che una costruzione è utilizzata, anche i suoi elementi costitutivi costanti ne sono rafforzati.

Le costruzioni sono però costituite anche da slot schematici che possono essere riempiti da elementi lessicali differenti, la cui presenza all'interno del pattern è soggetta a vincoli specifici imposti dalla costruzione che riguardano il tipo e la forma del *filler*. Il numero dei diversi *items*, che sono istanza di uno slot costituisce la misura della *frequenza type* della costruzione. Numerosi studi hanno mostrato come quest'ultima sia un fattore di fondamentale importanza nel determinare la predisposizione di un pattern ad accogliere al suo interno nuovo materiale lessicale (cfr. Cap.2, par. 2.3.1)

Uno degli obiettivi dei modelli usage-based è quello di fornire una spiegazione, oltre che degli aspetti regolari e predicibili di una lingua, anche dei fenomeni idiosincratici e *language-specific*.

Centrale nei modelli usage-based è la teoria degli esemplari, elaborata in psicologia cognitiva, che si basa sull'assunto che gli esseri umani possiedano la capacità di trattenere nella memoria dettagli sensoriali legati ad esperienze idiosincriche di interazione con l'ambiente circostante. A livello linguistico gli esemplari si presentano sotto forme diverse, da singoli segmenti, come le vocali, a interi sintagmi, raggruppati insieme seguendo i criteri di similarità e frequenza. La loro rappresentazione è molto importante perché riflette il modo in cui emergono le strutture linguistiche nell'interazione fra la cognizione umana e l'esperienza col mondo.

As words, phrases, or constructions are used they are mapped by speakers and hearers onto identical existing representations if they are present, strengthening them. Such representations are called EXEMPLARS and the term TOKEN is reserved for the actual instances found in usage. For the purpose of this mapping, identity involves both semantic and phonological features. If no identical exemplars are available the token is evaluated for its degree of similarity to other exemplars and represented as (metaphorically) close or distant to other exemplars. In this way clusters of related items are build up. (Bybee e Eddington, 2006:325)

La rappresentazione cognitiva della lingua, implica un fenomeno di categorizzazione a diversi livelli di astrazione, attraverso cui, il parlante attua un processo di astrazione sopra le singole istanze e sviluppa porzioni schematiche nelle costruzioni basandosi probabilisticamente sulla loro similarità e frequenza.

Si prenda come esempio un pattern del tipo *mangiare+N* (V+N), esso costituisce, nella mente del parlante, un unico esemplare di costruzione nonostante sia costituito da due parti ed ognuna di esse abbia una propria rappresentazione individuale. Se si prendono, invece, in considerazione due espressioni come *mangiare un gelato* e *mangiare un panino*, notiamo come esse siano costituite da una parte che è comune ad entrambe (in questo caso il verbo *mangiare*) e una parte che varia (il complemento oggetto diretto). Secondo una prospettiva usage-based, ogni volta che il parlante utilizzerà queste costruzioni, le parti comuni ad entrambe saranno proiettate sulla stessa rappresentazione, rafforzandola, mentre quelle variabili, andranno a costituire una categoria di elementi, insieme ad altri esemplari, che possono ricorrere con quel verbo. Il grado di generalità di tale categoria dipende dal numero e tipo di elementi con cui viene osservato *mangiare*.

È stato notato tuttavia come i soli esemplari non siano in grado di spiegare i processi di generalizzazione che riguardano sia la cognizione in generale che, nello specifico, il linguaggio. I modelli di rappresentazione basati solo sugli esemplari “fail to explain how exactly items cohere as a category” (Goldberg 2006:47). Per questo motivo, le teorie più recenti condividono l’idea che nel linguaggio, conoscenza *item-specific* e generalizzazioni coesistano e interagiscano.

Una delle caratteristiche principali di questi modelli è l’orientamento *bottom up* e *item-based*. I processi di schematizzazione e generalizzazione acquistano un ruolo di primaria importanza all’interno degli studi di acquisizione, ma anche in prospettiva diacronica e sincronica.

Molti studiosi si pongono come obiettivo primario proprio quello di indagare la natura di tale fenomeno, il suo ruolo nel processo di acquisizione linguistica da parte dei bambini ma anche nella creatività con cui gli adulti estendono pattern preesistenti nella lingua a nuovi usi in base alle esigenze comunicative.

L’importanza attribuita al fenomeno della generalizzazione deriva dall’osservazione che le lingue costituiscono sistemi molto regolari e che la nostra conoscenza delle strutture linguistiche forma una rete ben integrata.

We do not store an unlimited number of complete utterance representations; rather what we retain are instances at some level of abstraction. That is, we do not passively retain huge mental corpus, consisting of all the strings we have ever heard, as a computer

might do. Instead we constantly parcel out meaning, from abstractions, and generalize over instances we hear. (Goldberg, 2006:62)

1.3.1 Studi a sostegno dei modelli usage-based del linguaggio.

Numerosi sono gli studi, in vari ambiti linguistici, che avvalorano i principi alla base dei modelli usage-based.

È stato notato che nel processo di acquisizione del linguaggio vi sia una stretta corrispondenza fra l'input a cui è sottoposto l'apprendente e l'output che produce: i primi verbi ad essere appresi e utilizzati sono infatti quelli utilizzati più frequentemente dalle madri o dai *caregivers*. Tomasello (1992) ha notato come le prime rappresentazioni sintattiche dei bambini consistano in un insieme di strutture argomentali lessicalmente specificate (le cosiddette "isole verbali") costituite da un verbo che viene mantenuto stabile e da elementi nominali che variano all'interno del frame (*I wanna X, Open X*). Questi pattern costituiscono una sorta di punto di appoggio a partire dal quale i bambini giungono poi a costruzioni più complesse e astratte attraverso processi di analogia, schematizzazione e analisi distribuzionale.

Goldberg, Casenhiser, Sethuraman (2004) in uno studio basato sui corpora, hanno mostrato una correlazione fra la frequenza di occorrenza dei verbi all'interno di tre tipi di costruzione (VL, VOL, VOO) utilizzate dai bambini e quella degli stessi verbi nel parlato delle loro madri. I risultati mostrano che all'interno di ciascuna costruzione vi è un verbo che spicca per frequenza rispetto agli altri (per esempio, *go*, all'interno del pattern VL, il 54% delle occorrenze totali, mentre *get* ricorre con una frequenza pari al 6%; stessa cosa per il pattern VOL dove spicca *put* (31%) e per la costruzione VOO con *give*). Il fatto interessante è che una tendenza analoga si riscontra anche nel parlato che le madri rivolgono ai bambini: l'uso di una costruzione è dominato dalla ricorrenza, al suo interno, di un singolo verbo.

La genericità di significato di tali verbi e l'elevata frequenza con cui ricorrono nel linguaggio dei bambini li guida nel processo di generalizzazione del pattern a partire dall'input e facilita l'associazione fra il significato del verbo e quello della costruzione stessa: "highly frequent verb serves as a readily available prototype with which other verbs may be associated" (Goldberg et al. 2004:304).

In Casenhiser e Goldberg (2005) la tendenza riscontrata nello studio precedente, è stata verificata sperimentalmente direttamente su un gruppo di bambini dai 5 ai 7 anni. I risultati suggeriscono una maggiore facilità nell'apprendere nuove costruzioni al cui interno ricorrono verbi che nell'input compaiono con frequenza marcatamente elevata. Una tendenza analoga si riscontra anche in task di categorizzazione non linguistica (Homa, Dunbar, Nohre 1991, Markman e Maddox 2003), facendo dunque presupporre che tale meccanismo di apprendimento non sia esclusivo del linguaggio.

Since natural linguistic input tends to be skewed in this way, it seems that the associations of form and meaning that exist in languages do not need to be hard wired or universal –children are quite expert at learning the mappings, given general categorization strategies. (Casenhiser e Goldberg, 2005: 506)

Bannard e Matthews (2008), hanno indagato l'effetto che svolgono similarità e frequenza nel guidare l'abilità e la precisione dei bambini a ripetere sequenze di parole molto frequenti (*sit in your chair*) e varianti di tali sequenze meno comuni (*sit in your truck*).

Gli esiti dell'esperimento hanno mostrato come i soggetti abbiano una maggiore facilità nel ripetere stringhe di parole molto frequenti rispetto a quelle meno frequenti e impieghino un lasso di tempo significativamente minore nel ripetere le prime tre parole di una stringa quando esse costituiscono parte di un *chunk* (erano più veloci nel ripetere la sequenza *sit in your* quando essa era seguita da una parola come *chair* piuttosto che da una parola come *truck*).

Tali risultati sono in linea con una prospettiva usage-based per cui, durante il processo di apprendimento del linguaggio, il bambino inizialmente apprenderebbe un insieme limitato di espressioni direttamente dall'esperienza grazie ad abilità cognitive generali d'imitazione e *intention-reading* ed estenderebbe, in un secondo momento, tale conoscenza attraverso meccanismi di generalizzazione e analogia.

Theakston (2004) ha indagato il ruolo svolto dall'*entrenchment* nel contrastare gli errori di iper-generalizzazione da parte dei bambini e nel guidare gli adulti a fornire giudizi di accettabilità rispetto a determinate espressioni. Nel primo esperimento, ai soggetti, due gruppi di bambini rispettivamente di 5 e 8 anni, sono state sottoposte costruzioni argomentali grammaticalmente scorrette al cui interno ricorrevano verbi ad elevata frequenza (*Don't laugh me*) e verbi a bassa frequenza (*I dribbled teddy with water*).

Utilizzando una metodologia di gioco, è stato chiesto ai partecipanti di valutare l'accettabilità di ciascuna frase. I risultati mostrano come i bambini, nel fornire i loro giudizi, siano influenzati dalla frequenza di ricorrenza del verbo, poiché erano portati a ritenere più accettabili strutture argomentali non grammaticali quando al loro interno ricorrevano verbi a bassa frequenza.

È stato condotto un secondo esperimento, con lo stesso design ma con stimoli in parte differenti, per verificare se effetti simili di frequenza potessero essere osservati anche nei giudizi dei parlanti adulti. I risultati pur essendo in linea con l'esperimento condotto sui bambini, mostrano che per gli adulti non vi è una chiara relazione fra la frequenza dei singoli verbi e i relativi giudizi di grammaticalità. Secondo la studiosa, questo fatto potrebbe riflettere una complessa interazione fra la frequenza complessiva dei singoli verbi e la frequenza di utilizzo all'interno di costruzioni particolari.

In entrambi i casi lo studio sembra dimostrare che parlanti adulti e bambini siano sensibili alle proprietà distribuzionali dell'input: grazie ad esse, sono in grado di generalizzare le strutture argomentali conosciute ma allo stesso tempo ne vincolano l'iper-generalizzazione. Inoltre, emerge un'interessante riflessione circa il ruolo svolto dalle classi semantiche nella determinazione delle strutture argomentali in cui può comparire un verbo.

Knowing the semantic class to which a verb belongs, even for adult speakers, is not the only process involved in determining the argument structures in which that verb may appear. There is clearly an interaction between verb frequency and the degree to which a given verb's argument structure is governed by the semantic class to which it belongs. This suggests that narrowly defined semantic classes may be organized around prototypes based on the frequency of use of individual verbs rather than reflecting absolute membership criteria.

(Theakston, 2004:30)

La conoscenza linguistica di parlanti adulti e bambini dunque riflette la frequenza con cui questi ultimi utilizzano gli elementi lessicali e le strutture linguistiche.

From this perspective, it is the gradual building up of type and token frequencies that leads to abstraction beyond individual lexically-based constructions, but that also ultimately constrains productivity. (Theakston, 2004:32)

Altrettanto interessante è lo studio compiuto da Ambridge et al. (2008) i quali indagano il contributo e l'interazione fra *entrenchment* e classi semantiche nel fenomeno della generalizzazione.

Per verificare il rispettivo contributo dei due fattori e per individuare un eventuale meccanismo di interazione, Ambridge et al. hanno raccolto giudizi di accettabilità da parte di bambini e adulti, riguardo a frasi contenenti usi grammaticali e non grammaticali di verbi ad alta e bassa frequenza e verbi nuovi, appartenenti a tre classi semantiche differenti (*laughing, disappearing, falling*).

I risultati dell'esperimento mostrano come vi sia una tendenza da parte dei soggetti ad accettare più facilmente usi non grammaticali di verbi nuovi o a bassa frequenza rispetto a verbi ad elevata frequenza. Inoltre, sembra esserci un'effettiva interazione fra la formazione di classi semantiche verbali e il ruolo svolto dalla frequenza nel vincolare il fenomeno della generalizzazione delle strutture argomentali (cfr. anche Ambridge et al. 2012 e Ambridge et al. 2014).

Sempre sullo stesso tema si colloca lo studio di Boyd e Goldberg (2012) che hanno verificato la capacità di generalizzazione di tre gruppi di soggetti appartenenti a fasce di età differenti: un gruppo di bambini di 5 anni, un gruppo di bambini di 7 anni e infine un gruppo di adulti. Sono state create varie istanze di una costruzione del tipo NP₁NP₂V indicante un evento di avvicinamento o approccio, con NP₁ agente, NP₂ goal, e V un verbo inventato: *the doctor the construction worker feigos*.

Dopo una breve periodo di addestramento (*exposure trial*) durante il quale i soggetti hanno familiarizzato con varie istanze di nuove costruzioni frasali, questi ultimi sono stati sottoposti ad un task di comprensione a scelta forzata (*forced-choice comprehension task*) nel quale gli sono stati presentati anche altri tipi di costruzione e gli è stato chiesto di scegliere, dopo aver udito una frase, quale fra due filmati simultanei rappresentasse il significato espresso dalla frase appena udita. Per valutare il grado di astrazione delle rappresentazioni acquisite è stato variato il grado di *overlap* lessicale che le espressioni del test avevano con quelle dell'*exposure trial*. I risultati dell'esperimento, in linea con le teorie di apprendimento item-based, mostrano che in generale i bambini hanno una capacità di generalizzazione minore rispetto agli adulti, pur essendo stati tutti e tre i gruppi sottoposti allo stesso input. Però, se da una parte il gruppo di bambini più grandi si dimostra in ogni caso in grado di compiere tali generalizzazioni, il gruppo di bambini più piccoli risulta essere sensibile solo

all'*overlap* lessicale. Gli studiosi concludono che i bambini inizialmente hanno un atteggiamento conservatore e tendono a basarsi quasi esclusivamente sull'input ricevuto nel risolvere il task; questo esito è dovuto ad un mancato riconoscimento del pattern astratto esemplificato dalla nuova costruzione ma non ad una mancanza di esposizione alla costruzione stessa.

Recentemente Perek e Goldberg (2015) hanno condotto uno studio interessante riguardo al fenomeno della generalizzazione e al ruolo svolto dall'input al quale sono sottoposti i parlanti. Lo scopo dello studio era quello di indagare su quali basi i parlanti generalizzano una costruzione al di là dell'input al quale sono sottoposti, nella fattispecie, le parole che hanno esperito al loro interno.

I risultati mostrano che i parlanti non formulano generalizzazioni “alla cieca” sulla sola base statistica dell'input a cui sono sottoposti ma prendono in considerazione in modo critico anche la funzione della costruzione per verificare che possa essere estesa a nuove istanze. In particolare è stato notato che i partecipanti si comportano in modo conservatore quando non vi sono conflitti fra le restrizioni a cui sono soggetti i verbi e la funzione della costruzione di cui possono essere istanza, mentre hanno mostrato una maggiore predisposizione ad utilizzare i verbi in modo creativo quando tali estensioni sono più adatte alle esigenze comunicative (cfr. anche Wannacott 2011, Wannacott et al. 2008, Wannacott et al. 2012).

Ciò che accomuna questi studi è il fatto che le rappresentazioni evocate dai parlanti per risolvere i vari task sono astratte dall'esperienza linguistica stessa, con una diminuzione nella sicurezza nel fornire la propria risposta quando l'evidenza fornita dall'input è più limitata.

1.4 Sintesi.

Per riassumere, abbiamo visto come con il termine *usage-based* si intenda una famiglia di modelli che condividono una serie di assunti fondamentali, prima fra tutte la stretta relazione tra le strutture linguistiche e gli esempi d'uso della lingua. Il sistema linguistico del parlante è infatti saldamente ancorato agli “*usage-events*”, produzioni linguistiche concrete a partire dalle quali viene derivato il sistema di costruzioni linguistiche; essi non sono solo il prodotto del sistema linguistico dei parlanti, ma ne costituiscono anche l'input. Quest'ultimo svolge una funzione basilare soprattutto

nell'apprendimento della lingua da parte dei bambini e al quale in seguito si affiancano capacità di generalizzazione e astrazione.

In tutti i modelli, la frequenza riveste un ruolo di primaria importanza: un'elevata frequenza, in seguito al processo di automatizzazione nella produzione, provoca una riduzione fonologica; determina un mantenimento delle strutture, per cui le forme irregolari, che compaiono nella lingua più frequentemente, mantengono la loro irregolarità in misura maggiore rispetto alle forme più rare. Un'elevata frequenza può inoltre determinare un processo di grammaticalizzazione con la conseguente creazione di morfemi grammaticali e di cambiamenti nella struttura argomentale.

Una delle caratteristiche principali dei modelli *usage-based* è che esse prendono in esame non solo la competenza del parlante ma anche i processi di produzione e percezione che hanno luogo nell'utilizzo del linguaggio e per farlo raccolgono sia dati sperimentali che dati provenienti da grossi corpora rappresentativi del linguaggio naturale; questi ultimi rappresentano una risorsa molto importante per lo studio del linguaggio umano poiché restituiscono informazioni fondamentali circa la natura e della frequenza con cui i parlanti utilizzano le strutture linguistiche.

L'approccio basato sull'uso, in conclusione, si presenta come un modello in grado di spiegare adeguatamente determinati fenomeni linguistici senza ricorrere necessariamente al concetto di categoria innata. Vedremo poi, nel prossimo capitolo, come tale modello costituisca uno strumento importante per affrontare il tema della produttività sintattica.

2. Produttività sintattica: lo stato dell'arte.

2.1 Le nozioni di produttività, creatività e analogia fra generativismo e modelli usage-based del linguaggio.

La nozione di produttività linguistica costituisce un concetto complesso e sfaccettato che nel corso degli anni ha ricevuto molteplici definizioni spesso opposte fra loro, alcune dal carattere più tecnico, altre che invece riflettono usi più laschi e informali del termine.

Tutte però si basano su due osservazioni: il fatto che la lingua evolve nel corso del tempo fino a modificare anche gli aspetti più profondi e strutturali del sistema e il fatto che i parlanti utilizzano le strutture linguistiche in modo creativo per rispondere a sempre nuove esigenze comunicative.

Conoscere una lingua significa essere in grado di comprendere e produrre un numero virtualmente illimitato di espressioni linguistiche nuove.

Per spiegare la natura di tale conoscenza, la grammatica generativa definisce il linguaggio come un codice formale, un insieme di regole che stabiliscono quali strutture linguistiche sono accettabili e quali no all'interno di una determinata lingua, andando a delineare la competenza (*competence*) linguistica del parlante.

Conoscere una lingua significa dunque conoscere quali sono le regole che la governano. Chomsky, che pone al centro dell'indagine sintattica la frase, ritiene che il compito della grammatica sia proprio quello di individuare quali sono le sequenze di parole accettabili all'interno di una lingua e quali siano invece da considerare non grammaticali.

The fundamental aim in the linguistic analysis of a language L is to separate the grammatical sequences which are the sentences of L from the ungrammatical sequences which are not sentences of L and to study the structure of the grammatical sentences.
(Chomsky, 1957:13)

Gli approcci generativi, adottano dunque una prospettiva riduzionistica, in virtù della quale, la grammatica, composta da unità minime, attraverso meccanismi sintattici di tipo computazionale e compositazionale è in grado di derivare un numero illimitato di

combinazioni. La frase costituisce dunque una configurazione prestabilita, ottenuta tramite l'applicazione di regole standard e interpretabile sul piano semantico in modo compositivo.

Un comportamento linguistico produttivo, secondo tale visione, è un comportamento linguistico regolare, ed è il risultato dell'applicazione di regole grammaticali memorizzate nella mente del parlante.

The theory that regular forms are generated by rule and irregular forms are retrieved by rote is pleasing not only because it explains the differences in productivity between the two patterns but also because it fits nicely into the larger picture of the design of language. (Pinker, 2015:17).

Ciò che emerge è dunque una prospettiva binaria e categorica, legata a modelli algebrici della grammatica in cui si distingue ciò che è produttivo da ciò che non lo è, e in cui un insieme numericamente limitato di regole opera su un lessico statico, insieme al quale forma un sistema generativo. L'estensione delle regole a nuovi casi è accettata purché vengano rispettate le condizioni imposte dalle regole stesse; le eccezioni, i casi irregolari, sono invece catalogati nel lessico che costituisce un modulo linguistico distinto. In questo modo si delineano, all'interno del dominio della grammatica, due moduli separati, la *core grammar*, costituita da un insieme ben delimitato di regole grammaticali e una *periferia linguistica*. Secondo tale prospettiva, il lessico costituisce dunque una sorta di appendice della grammatica nella quale vengono elencate le varie irregolarità della lingua.

All'interno di tale prospettiva, gli studiosi hanno spesso cercato di tracciare una linea di separazione tra il concetto di *creatività* linguistica e quello di *produttività*.

Il primo, di carattere più simile all'analogia, si ricollega ad un cambiamento linguistico che va al di là delle regole imposte dalla lingua e ha come output la formazione di nuovi pattern altrimenti considerati non produttivi; il secondo invece consiste in un processo regolare.

Productivity can be distinguished from creativity, although it is hard to draw a consistent line between the two. It may be the case that productivity can be seen as **rule-governed**, and creativity seen as **rule-changing** and equated with the use of analogy, but this is not settled. (Bauer, 2001: 98)

Tuttavia, una netta distinzione fra lessico e sintassi, non permette di classificare determinati fenomeni linguistici per i quali sarebbe necessaria una conoscenza speciale, “un’appendice alla grammatica” per rendere conto di ciò che quest’ultima, insieme al lessico e ai principi di semantica compositiva non è in grado di spiegare. In particolare, non è chiaro se le costruzioni idiomatiche e certi pattern semi-produttivi, debbano essere catalogate nel lessico o possano essere spiegate attraverso l’applicazione di regole sintattiche. Tale repertorio linguistico, onnipresente e comune a tutte le lingue, non può d’altronde essere considerato come accessorio alla conoscenza linguistica proprio a causa dell’ingente quantità di fenomeni che vi rientrano e dall’alto grado d’interazione che essi hanno con la grammatica.

Recentemente, grazie al contributo della linguistica dei corpora, numerosi studi hanno mostrato come il concetto di produttività sia più ampio e non si limiti alla sola competenza linguistica del parlante. Il cambiamento linguistico e la sua evoluzione nel corso degli anni non sono dovuti esclusivamente all’applicazione di regole grammaticali ma hanno origine nell’utilizzo e nell’esperienza che i parlanti fanno con la lingua.

La produzione linguistica e il processo di selezione sintattica non sono considerati come processi monolitici ma, al contrario, come meccanismi complessi d’interazione fra innovazione e ripetizione; il riconoscimento di questo aspetto è fondamentale per comprendere la facilità con cui i parlanti, adulti e bambini, acquisiscono e fanno uso della lingua.

Speakers are at once impressively creative and impressively repetitive. The creative aspect of language allows speakers to express new ideas in new contexts. [...] The formulaic aspect presumably simplifies the speakers’ task, in that many utterances or parts of utterances can be pulled directly off the mental shelf without having to build every utterance anew, from scratch.

(Goldberg, in Hoffman, & Trousdale, 2013: 26)

Il riconoscimento di questa duplice natura ha portato verso un crescente consenso nei confronti delle cosiddette teorie *usage-based*, modelli di analisi del linguaggio che ritengono che le strutture linguistiche emergano attraverso l’uso e l’esperienza con la lingua, e che la rappresentazione cognitiva della lingua stessa nella mente del parlante sia dinamica e pronta al cambiamento. Tali modelli si distinguono dagli approcci generativi, poiché ritengono che la lingua sia un inventario di unità convenzionali che vanno dalla piena generalità alla completa idiosincrasia; la competenza del parlante,

comprende la memorizzazione di regole generali (schemi) che, attraverso una prospettiva bottom-up, sono create a partire dall'esperienza che il parlante ha con esemplari specifici.

All'interno della lingua operano due meccanismi. Da una parte un effetto di *conservazione*, quello che Israel (1996) chiama *production principle* ("utterances should sound like things the speaker has heard before"), che vincola le generalizzazioni che i parlanti sarebbero pronti ad estrarre dall'input e riflette il fatto che "people tend to talk like the people they identify with" (Israel, 1996:228) dall'altra, un meccanismo di stimolazione che al contrario, incoraggia quindi il parlante ad estrarre le generalizzazioni e ad andare oltre lo status quo, estendendo i pattern conosciuti; a tale riguardo Israel parla di *comprehension principle* che rappresenta la forza dell'innovazione, ed è dovuto al fatto che "people will seek to accommodate and make sense of even the most unexpected novel utterances" (id.:228)

Questa duplice pressione spinge il parlante a rispettare le convenzioni imposte dalla propria lingua, ma allo stesso tempo a trovare nuove soluzioni per rispondere a sempre nuove esigenze comunicative.

In questo modo, la distinzione fra i concetti di creatività, analogia e produttività come riflesso di una discriminazione fra processi eccezionali e processi regolari e sistematici è rifiutata a favore dell'immagine della produttività come *continuum* di fenomeni. Il fatto stesso di porre al centro di indagine la nozione di costruzione, permette di andare oltre alla distinzione fra elementi lessicali e strutture frasali che al contrario, sono collocati lungo un unico continuum lessico-sintassi.

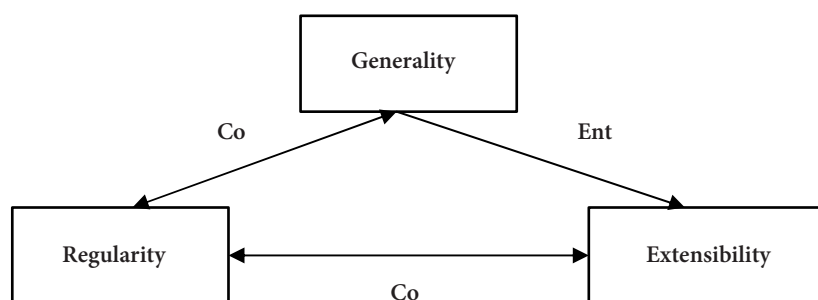
Secondo tale prospettiva, la produttività si manifesta sotto tre aspetti (Barðdal 2008): come estensione del concetto di *regolarità*, come realizzazione del concetto di *estendibilità* e come quello di *generalità*, legati fra loro da relazioni di co-occorrenza e implicazione.

Abbiamo visto come la nozione di *regolarità* sia alla base della definizione che della produttività è stata data dalle teorie di stampo generativista, e consista nell'abilità dei parlanti di comprendere e generare frasi mai incontrate prima. Secondo tale prospettiva, un comportamento produttivo è un comportamento regolare, ed è l'output dell'applicazione di una regola grammaticale (rule-based); un pattern regolare è anche semanticamente trasparente e scomponibile, se così non fosse, costituirebbe una categoria fissa e immodificabile. In questo senso il concetto di regolarità è collegato anche a quello di composizionalità.

Con il termine *generalità* si intende la capacità di un pattern di essere schematicamente aperto, la mancanza di restrizioni alla sua applicabilità ma anche il suo essere di “default” e di ampia copertura. Esso si ricollega anche alla nozione di schematicità, per cui il livello più astratto di un pattern è anche quello più schematico e più generale, quindi sottoposto a un minor numero di restrizioni.

Infine il concetto di *estendibilità* si riferisce, come abbiamo visto sopra alla flessibilità sintattica di un pattern, ovvero alla sua capacità di accettare al suo interno nuovi elementi lessicali. Tale concetto si può inoltre riferire alla predisposizione che una determinata costruzione o categoria linguistica ha, di sviluppare nuove funzioni pragmatiche e semantiche.

Figura 2 I tre sotto-concetti di produttività secondo Barðal (2008: 23).



La relazione che lega i concetti di generalità ed estendibilità è di tipo implicativo (*Entailment*): secondo Barðdal, una categoria generale è sempre estendibile ma non è vero il contrario.

I concetti di generalità-regolarità e regolarità-estendibilità sono invece legati da un legame di co-occorrenza, nel senso che la presenza di uno non implica necessariamente la presenza dell'altro, anche se, in molte lingue, per numerose categorie, tali concetti sono spesso concomitanti.

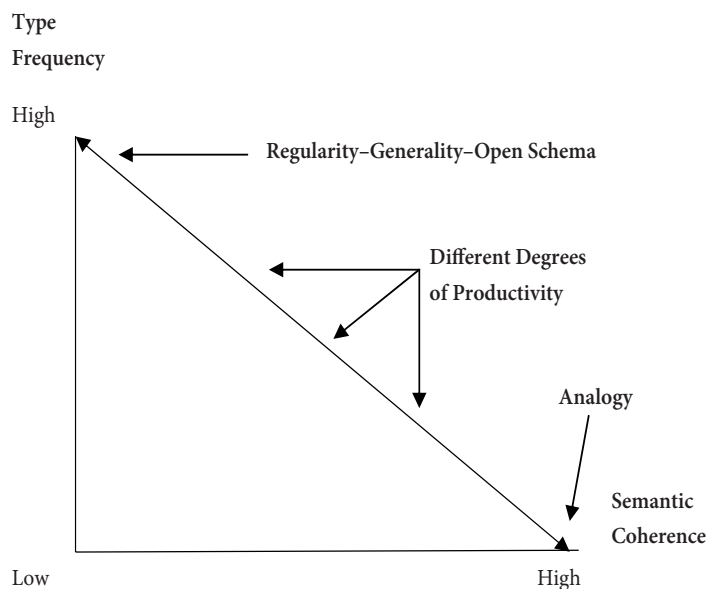
Secondo Barðdal il concetto di estendibilità tuttavia mantiene a livello teorico un'importanza primaria mentre quelli di regolarità e di estendibilità, pur essendo ad esso strettamente collegati, sono secondari e derivati.

È proprio a causa dell'intricato rapporto che intercorre fra tali nozioni che il fenomeno della produttività costituisce un concetto complesso e di difficile definizione.

La produttività è descritta come un gradiente (*cline*) che vede alle due estremità, da una parte, gli schemi più aperti, generali e regolari e dall'altra le estensioni analogiche

basate su un unico esemplare specifico, mentre, al centro del continuum si trovano categorie caratterizzate da gradi diversi di produttività.

Figura 3 Aspetti differenti del *cline* di produttività secondo Barðal (2008:38)



In questo modo, la discussione si sposta da un piano di carattere qualitativo ad uno di carattere quantitativo con l'obiettivo preciso di quantificare il fenomeno della produttività, di individuarne i fattori stimolanti e quelli vincolanti. La visione strettamente categoriale e binomiale è sostituita da un'immagine scalare del fenomeno, considerato una variabile ordinale.

Il fenomeno della produttività si manifesta all'interno di molteplici aree dell'uso della lingua. Durante il processo di acquisizione del linguaggio, guida i bambini a sviluppare l'abilità a generalizzare al di là dell'input, formando schemi e sperimentando nuove combinazioni di parole (cfr. Bowerman 1988, Pinker 1989, Tomasello 2003). Usi creativi della lingua sono compiuti anche dai parlanti adulti che sperimentano nuove espressioni non convenzionali per rispondere a nuove esigenze comunicative: il mondo, la società, i bisogni cambiano e i parlanti riflettono tali cambiamenti nella lingua, allargando le costruzioni a nuovi usi e creando nuove espressioni ex-novo.

La produttività linguistica può essere analizzata attraverso una duplice prospettiva: una di tipo diacronico (cfr. Perek 2016, Barðdal 2008), che studia l'evolvere della lingua, il suo cambiamento nel corso del tempo, e quindi il modo in cui le costruzioni accettano al loro interno nuovi elementi lessicali, ed una di tipo sincronico (cfr. Zeschel 2012) che,

soffermandosi su un preciso arco temporale, studia la variazione linguistica, ovvero quali sono le costruzioni attestate e in che misura tali pattern possono andare in contro a variazioni creative e ad estensioni e, allo stesso tempo, cerca di individuare i vincoli e le restrizioni semantiche che agiscono sugli elementi lessicali che possono esserne istanza.

2.2 Produttività: dalla morfologia alla sintassi.

2.2.1 Definizioni e misure di produttività morfologica

Lo studio della produttività, di stampo usage-based, nasce in morfologia e riguarda l'analisi dei processi di formazione e derivazione delle parole. In tale ambito, la produttività consiste nella possibilità per qualsiasi procedimento morfologico di poter essere utilizzato per la costituzione di nuove formazioni.

È stato notato come determinati processi morfologici siano più comuni di altri (sono rappresentati da più tokens all'interno dei dati provenienti dai corpora), abbiano una più ampia varietà di forme (numero di types) o una maggiore possibilità di dare adito a nuove formazioni.

Fra le numerose definizioni che sono state date alla produttività in morfologia ne riporto alcune fra le più importanti e accettate.

- i. La probabilità che un modello di formazione delle parole sia utilizzato come modello per la creazione di nuove parole. (Aronoff, 1983: 163)
- ii. La probabilità che un modello sia applicato nella creazione di nuove forme. (Bybee, 1995: 430)
- iii. L'interazione fra la potenzialità di un processo morfologico a generare forme ripetitive e non creative, e la misura in cui esso è utilizzato, invece, per creare nuovi elementi lessicali. (Bauer, 2001:211)

Tutte riconoscono nella produttività, il concetto di estendibilità, ovvero la possibilità che un determinato processo possa essere esteso a nuove parole e il collegamento fra tale possibilità e l'esperienza linguistica del parlante.

I vari approcci alla produttività morfologica hanno in comune il fatto di averne

sottolineano la natura scalare: i diversi procedimenti morfologici possono essere considerati produttivi in misure diverse. Proprio per questo motivo, i dibattiti, più che trovare una definizione tecnica al concetto, si sono incentrati sulla ricerca di un modo per quantificare e misurare tale nozione.

Uno dei paradigmi più influenti, in tale ambito, risale al lavoro di Bayeen e colleghi (Bayeen e Lieber, 1991; Bayeen 1992, 2001, 2009) i quali hanno sviluppato una serie metodi per misurare la produttività basati sull'analisi statistica di corpora linguistici. La produttività è analizzata come la probabilità che un determinato procedimento morfologico dia luogo a una nuova formazione all'interno del lessico di una lingua. I vari metodi elaborati da Bayeen si basano prevalentemente su tre elementi: la *frequenza token* di una determinata categoria (denotata $N(C)$), ovvero il numero totale di ricorrenze di un determinato pattern all'interno di un corpus o di un testo, la *frequenza type* (denotata V), che indica il vocabolario di un verbo, ovvero il numero dei diversi elementi lessicali che sono istanza di uno slot argomentale, e gli eventi rari, in particolare gli *hapax legomena*, parole che compaiono solo una volta all'interno del corpus e che hanno, pertanto, la massima probabilità di essere neoformazioni.

Il metodo consiste nel calcolare la probabilità che l'insieme di tutte le forme costituite da un determinato procedimento morfologico possa essere espanso a nuovi tipi (*types*), generando neoformazioni. Questo indice di produttività è stato chiamato da Bayeen “*category-conditioned degree of productivity*” ed è ottenuto calcolando il rapporto fra il numero di hapax e il numero di tutti i tokens all'interno del corpus; esso predice che maggiore è la proporzione di hapax nel campione di tokens considerato, tanto più produttivo sarà il procedimento morfologico che li ha generati. Se invece il procedimento ha dato vita a molti types o a molte istanze, ma a poche neoformazioni, si tratterà di un procedimento scarsamente produttivo.

La formula per calcolare tale indice di produttività è:

$$P = VI/N(C)$$

Tale misurazione si basa sull'assunto euristico che un neologismo avrà occorrenza limitata all'interno di un corpus; è bene sottolineare come questa idea non rifletta sempre la reale distribuzione delle parole: spesso i parlanti coniano nuove parole per determinati bisogni comunicativi e le utilizzano in successione, in altri casi avviene che determinate espressioni compaiano in un corpus un'unica volta per caso, nonostante siano forme lessicalizzate e familiari alla maggior parte dei parlanti.

Il vantaggio legato a tali misurazioni risiede tuttavia nel fatto che forniscono strumenti

in grado di quantificare il potenziale d'innovazione di un determinato processo linguistico, fornendo informazioni circa la probabilità che esso sia esteso a nuovi membri, e al suo livello di saturazione, a partire dalla ricorrenza e dalla distribuzione delle istanze già attestate.

Altre misure sono state proposte, come il semplice calcolo della frequenza *type*. Tali metodi riflettono aspetti differenti della produttività e di come essa possa essere realizzata nei testi, nei dizionari e in corpora più ampi, ed è importante sottolineare come, secondo lo studioso, i vari metodi andrebbero utilizzati in parallelo poiché ognuno di essi mette in evidenza un aspetto particolare della produttività.

Sebbene i procedimenti illustrati da Bayeen siano stati ampiamente accettati, grazie anche alla loro applicabilità pratica, non tutti gli studiosi concordano sulla natura probabilistica della produttività morfologica.

Numerosi studi hanno poi cercato di individuare una correlazione fra le cause della produttività dei procedimenti morfologici e il modo in cui l'input viene rappresentato nella mente dei parlanti. Tali studi partono comunque dall'osservazione, portata avanti da Bayeen, che esiste una stretta correlazione fra frequenza e produttività.

Bybee (2010), adottando una prospettiva di stampo *usage-based*, ha analizzato il contributo della *frequenza type*. Quest'ultima è definita come il numero delle diverse forme di parola che sono istanze di uno schema particolare. Alla base della produttività è postulato un meccanismo di estensione analogica a partire da istanze specifiche; una costruzione caratterizzata da un'elevata frequenza *type*, avrà una probabilità maggiore di essere utilizzata per creare nuove formazioni, per il fatto che ha un maggior numero di candidati su cui basare l'analogia.

L'effetto della frequenza *type* può essere vincolato da determinati fattori. Per esempio le istanze di uno schema che hanno una *frequenza token* molto elevata contribuiscono alla produttività in misura minore poiché sono caratterizzate da una maggiore autonomia e dalla perdita di analizzabilità.

Questo avviene perché, se una forma morfologicamente complessa è molto frequente, la sua rappresentazione come unità si rafforza nel lessico mentale. Di conseguenza, il percorso di accesso diretto a quella forma diventa predominante durante l'elaborazione linguistica, senza il bisogno di scomporla in morfemi. Questo favorisce un'opacizzazione della forma complessa sia dal punto di vista morfofonologico che da quello morfo-semantic. Al contrario, una forma poco frequente, difficilmente ha una rappresentazione autonoma nel lessico mentale, per cui l'accesso avviene tramite

composizione morfologica. In questo modo sono i morfemi che la compongono ad essere attivati e dunque rafforzati nella mente del parlante, il che contribuisce alla loro produttività. A tale riguardo, Bybee (2010) cita il caso dei verbi irregolari inglesi. All'interno di tale categoria possiamo notare come i verbi a bassa frequenza mostrino una maggiore predisposizione a subire un processo di livellamento analogico e regolarizzazione sulla base del modello offerto dalle forme regolari: si pensi a verbi come *weep*, *creep*, *leap*, per i quali, la maggior parte dei vocabolari inglesi riporta sia la forma irregolare (*wept*, *crept*, *leapt*) che quella regolare (*weaped*, *creeped*, *leaped*). Al contrario, verbi a elevata frequenza come *keep* e *sleep*, che sono più radicate nella mente dei parlanti, mostrano una maggiore resistenza alla regolarizzazione e una predisposizione maggiore a mantenere la propria irregolarità nella flessione al passato (*kept*, *slept*).

This conservative behaviour of high frequency forms is related to the faster lexical access of high frequency forms: the more a form is used, the more its representation is strengthened, making it easier to access the next time. Words that are strong in memory and easy to access are not likely to be replaced by new forms created with the regular pattern. (Bybee e Thompson, 1997:380)

Un altro fattore importante nel determinare la produttività di uno schema è rappresentato dalla coerenza semantica e fonologica interna allo schema stesso. Ciò significa che deve esserci sufficiente somiglianza tra i types che contribuiscono al radicamento dello schema, affinché questo possa effettivamente formarsi.

Le importanti intuizioni di Bybee sono state riprese e applicate successivamente in ambito sintattico per analizzare il fenomeno della produttività di pattern più estesi.

2.2.2 Dalla morfologia alla sintassi

La prima definizione che è andata al di là del solo campo morfologico, risale a Hockett (1958), il quale afferma che:

The productivity of any pattern- derivational, inflectional, or syntactical- is the relative freedom with which speakers coin new grammatical forms by it. (Hockett, 1958:307)

Lo studioso analizza il fenomeno della produttività in vari ambiti fra cui la flessione plurale in inglese, con l'aggiunta del morfema *-s*, il fenomeno di derivazione attraverso l'aggiunta del suffisso *-ly*, e l'estensione delle frasi con copula del tipo "X è Y".

Più tardi anche Langacker fornirà una definizione che abbraccia ogni tipo di costruzione:

Productivity is a matter of how available a pattern is for sanction of novel expressions.
[...] patterns of comparable generality can easily differ in their degree of productivity.
(Langacker, 1999: 114)

La definizione di Langacker è importante perché parlando genericamente di *pattern*, sottolinea come la nozione di costruzione, intesa dalla grammatica costruzionista come corrispondenza di forma e significato a vari livelli di specificità, abbracci tutti i campi di indagine linguistica e sia alla base di un continuum lessico-sintassi.

È a partire dal concetto stesso di costruzione, come corrispondenza di forma e significato, che è dunque possibile tracciare un parallelismo fra la produttività morfologica e quella sintattica.

Barðdal (2008) sottolinea per esempio che, per quanto riguarda la struttura interna, esistono delle notevoli somiglianze fra le parole e le frasi, fra processi morfologici e processi sintattici.

Syntactic constructions like argument structure constructions are, in language use, always instantiated by some lexical material, for instance the predicate which occur in them. This parallels to morphological constructions or words which are instantiated by their morphemes. That is, argument structure constructions are syntactic patterns, instantiated by specific number of verbal types, exactly like word formation patterns are morphological patterns, instantiated by a certain number of morphological based types. (Barðdal, 2008: 30-31)

Il parallelismo compiuto da Barðdal riguarda da una parte i concetti di flessione e derivazione di stampo morfologico e dall'altra quelli di regolarità ed estendibilità che secondo la studiosa sono alla base della nozione di produttività sintattica¹. Barðdal individua un ulteriore processo che consiste nella *formazione* di nuovi pattern che la studiosa esemplifica facendo riferimento alla nascita di nuovi morfemi, come il suffisso inglese *-oholic* e di nuove costruzioni, come la *way-construction*. Barðdal (2008) viene così a delineare una divisione tripartita fra *token*, *type* e *patterns*, che si trovano in

¹ In morfologia, la *flessione* implica che siano prodotte nuove forme (o tokens) a partire da types già

entrambi i livelli di analisi linguistica.

Figura 4 Tabella ripresa e adattata da Barðdal (2008:32).

	Nuovi token di pattern esistenti	Nuovi type per pattern esistenti	Nuovi pattern
Morfologia	Nuove forme di parola	Nuovi elementi lessicali per specifici pattern morfologici.	<i>-ohoilc</i>
Sintassi	Frasi	Nuove entrate lessicali verbali per specifiche costruzioni a struttura argomentale.	<i>Way-construction</i>

Amir Zeldes (2012) ha realizzato, recentemente, uno studio volto ad indagare se le stesse tecniche impiegate per calcolare la produttività dei processi morfologici possano essere applicate agli slot argomentali delle costruzioni. In particolare, lo studioso ha adottato la stessa metodologia utilizzata da Bayeen (2001, 2009), nello studio della produttività degli affissi morfologici, per stimare la produttività delle costruzioni a struttura argomentale costituite da uno o più slot aperti. Il parallelismo tracciato da Zeldes è riassumibile all'interno di una tabella:

Figura 5 Riadattamento dello schema proposto da Zeldes (2012:25) che raffigura il parallelismo fra processi morfologici e sintattici.

	<i>Morfologia</i>	<i>Sintassi</i>
<i>Costante</i>	Processo (es. un affisso che necessita di una base)	Processo (es. una costruzione con uno slot aperto)
<i>Variabile</i>	Base morfologica	Slot

In particolare, Zeldes (2012) assume, come Barðdal, che a partire dal concetto di costruzione sia possibile mettere sullo stesso piano i processi di formazione di una parola e quelli di un pattern sintattico.

If we see word formation and phrase building as comparable constructions, the most straight-forward translation of morphological processes to syntactic processes is to equate an invariable recurrent part of the construction as indicative of the productive process (much like a morphological affix) and the variable part (a content lexeme filling some slot of the construction) as the item operated upon by the process (much like a morphological stem). (Zeldes, 2012:24)

Entrambi i processi constano di una parte invariabile, rappresentata dall'affisso in morfologia e dalla porzione lessicalmente specificata della costruzione in sintassi, e di una parte variabile, costituita dalla base morfologica in un caso e da uno o più slot aperti nelle costruzioni a struttura argomentale.

Si pensi al suffisso della prima coniugazione nella flessione dei verbi italiani, *-are*, essa costituisce la costante del processo morfologico, a cui possiamo aggiungere numerose variabili, o basi morfologiche (*cucin-*, *studi-*, *pens-*, ecc.), l'insieme delle quali andrà a costituire il vocabolario, ovvero la stima della frequenza type di quel preciso processo.

Allo stesso modo, in sintassi, se prendiamo in considerazione l'espressione *mangiare X*, l'elemento costante sarà costituito dal verbo mentre la parte variabile, dallo slot argomentale che può essere istanziato da vario materiale lessicale (*panino*, *pizza*, *mela*, *gelato*, ecc.) nel rispetto dei vincoli semantici imposti dalla costruzione.

Una caratteristica che differenzia le costruzioni a struttura argomentale dai pattern morfologici è che le prime possono avere molteplici slot; pertanto la scelta di quali di essi considerare determinanti nell'analisi della produttività varia a seconda del tipo di analisi che si vuole condurre e del tipo di costruzione da analizzare: se da una parte Barðdal, infatti propende ad analizzare l'estensione dello slot verbale a nuove entrate lessicali, lo studio di Zeldes verte al contrario sugli slot argomentali. La differenza tuttavia è marginale, entrambi gli approcci partono dall'idea che la costruzione sia caratterizzata da posizioni libere riempite da materiale lessicale.

2.2.3 Definizioni di produttività sintattica.

In ambito prettamente sintattico, il concetto di produttività è stato affrontato da numerosi studiosi, fra cui Goldberg (1995; 2006; 2011), Clausner e Croft (1997), Barðdal (2008), Zeldes (2012), Zeschel (2012), Perek (2015, 2016). I vari studi hanno in comune il fatto di aver adottato una prospettiva di carattere usage-based, e di aver inteso la produttività come un fenomeno scalare che abbraccia pattern linguistici pienamente produttivi, semi-produttivi e estensioni analogiche e, in particolare, come la capacità di estendere le costruzioni a nuovi elementi lessicali, mentre ciò che li distingue, come vedremo nel prossimo paragrafo, è la relativa importanza che ognuno di essi ha attribuito ai fattori che determinano la produttività sintattica e al modo in cui essi interagiscono fra loro.

Goldberg (1995:120) definisce la produttività delle costruzioni come “the ability to be extended to new hypothetical verb forms”; essa è dovuta alla tendenza dei parlanti a generalizzare le costruzioni al di là dei loro contesti originali. Con il termine *partial productivity*, essa sottolinea come le costruzioni però non possano essere estese a qualsiasi elemento lessicale ma siano soggette a vincoli di carattere semantico. La produttività, in linea con la prospettiva usage-based, è inoltre considerata un fenomeno scalare legato a una distinzione di carattere quantitativo più che qualitativo: è rappresentata da un gradiente che va da costruzioni poco produttive a quelle pienamente produttive.

Come esempio di costruzione poco produttiva, la studiosa propone il caso della *risultativa*, una costruzione caratterizzata da numerosi vincoli di natura semantica e di carattere morfofonologico che limitano il numero e il tipo di verbi che possono ricorrere al suo interno (cfr. Goldberg, 1995; Goldberg e Jackendoff, 2004 e Boas, 2005 per una visione alternativa).

Si vedano alcuni esempi, ripresi da Goldberg, 1995:

- 6) a. She shot him dead.
b. ?? She blasted him dead.

- 7) a. She cried herself to sleep.
b. ?? She wept herself to sleep.

Allo stesso tempo però, la costruzione non presenta i tratti della completa idiomaticità, Goldberg (1995), infatti, riporta casi in cui essa è utilizzata produttivamente:

- 8) Drive your engine clear.
- 9) She could wonder herself crazy over the human eyebrow. (Goldberg,1995:137)

Sul versante opposto del continuum, si colloca invece la “*way- construction*” caratterizzata da un elevato tasso di produttività e soggetta a vincoli semantici di carattere più generale che la portano ad accettare al suo interno un vasto numero di verbo. Ecco alcuni esempi proposti da Goldberg (1995), tratti da corpus Oxford University Press:

- 10) [They] hoped they too could massage their way to keeping power.
- 11) [...] Their customers snorted and injected their way to oblivion and sometimes died on the stairs.

Il fatto che al suo interno compaia una vasta e varia gamma di verbi la rende più aperta ad accettare anche materiale lessicale nuovo.

Goldberg inoltre, traccia un’interessante correlazione fra il fenomeno della produttività e quello della *coercizione* per cui “a construction coerces the meaning of a word so that the word is construed to be compatible with the construction’s function” (id.:159).

Prendiamo un esempio:

- 12) Sam squeezed the rubber ball inside the jar. (Goldberg, 1995:158).

In (12), l’idea del movimento veicolata dalla frase non è codificata dal verbo *to squeeze* (infilare, comprimere) o dalla preposizione *inside* (dentro) né dalla combinazione dei due elementi. A tale riguardo Goldberg in linea con Talmy (1977) e Carter (1988) parla di un processo di adattamento (*accomodation*) per cui è la costruzione che porta determinate parole ad assumere un significato che sia compatibile con la funzione della costruzione stessa.

Affinché tale processo si realizzi però, è necessario che vi sia una relazione fra il significato inerente della parola che subisce tale fenomeno di adattamento e la sua interpretazione *forzata*.

Coercizione e produttività secondo Goldberg si riferiscono ad aspetti differenti dello stesso fenomeno, riducendo la distinzione ad un piano prettamente terminologico.

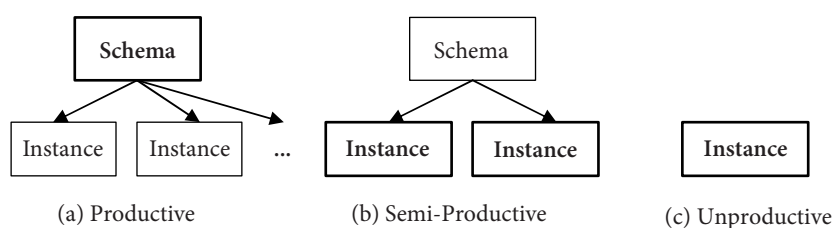
A construction is considered to be productive to the extent that it can coerce new words to appear in it. If there is a difference, it is that *coercion* tends to be used when there is an intuition that a word is construed somewhat unusually in order to appear in the construction (i.e., the construction “coerces” the word). [...] but the dividing line is not clear cut. (Goldberg et al., 2011: 1238)

Clausner e Croft (1997) in uno studio sulla produttività e la schematicità delle metafore, in cui tracciano un parallelismo fra il processo di derivazione di una forma linguistica e quello di estensione di una categoria semantica, hanno individuato una corrispondenza fra il grado di radicamento di una costruzione e il suo livello di produttività. Quest’ultima è definita come “the degree to which a schema is more entrenched than its instantiations” (Clausner e Croft, 1997: 255).

In particolare, la produttività si presenta come un continuum graduato lungo il quale si possono individuare tre punti cruciali: fenomeni ad elevata produttività, fenomeni semi-produttivi e fenomeni non produttivi a cui gli studiosi fanno corrispondere tre tipologie di metafore, le metafore convenzionali (*good is up*), gli idiomi trasparenti (*spill the beans*) e gli idiomi opachi (*kick the bucket*).

Ciascun livello di produttività è caratterizzato da un differente livello di radicamento dello schema e delle sue istanze. Le costruzioni più produttive sono quelle caratterizzate da uno schema ben radicato a partire dal quale è possibile generare nuove istanze, i pattern semi-produttivi presentano uno schema meno radicato ma istanze specifiche ben radicate, infine, per le metafore non produttive, non è possibile postulare l’esistenza di alcuno schema sovraordinato, per questo motivo, sono rappresentate da isole lessicalmente specificate e ben radicate nella mente del parlante.

Figura 6 Rappresentazione della produttività secondo Clausner e Croft (1997).



Per gli studiosi, ciò che determina il grado di produttività di una costruzione è “the proportion of a schema’s range which can be instantiated as expressions [i.e., the range of concepts consistent with the schema] which is actually manifested” (id.:257).

Più tardi, Barðdal (2008), in linea con Goldberg, identifica la produttività delle costruzioni argomentali con il concetto di *estendibilità*, ovvero con la capacità delle costruzioni di essere estese a nuovi elementi lessicali.

Come abbiamo visto sopra, la studiosa ritiene che la nozione di produttività come estendibilità di un pattern, interagisca con altri due concetti, quelli di *generalità* e *regolarità*, da essa derivabili, e che sono collegati fra loro da relazioni di co-occorrenza e/o di implicazione.

In questo modo, la definizione della produttività sintattica delineata da Barðdal (2008) costituisce, in un certo senso, la sintesi fra due fenomeni: da una parte l’abilità dei parlanti nel generare e comprendere espressioni e frasi mai incontrate in precedenza, dall’altra la capacità di un determinato pattern sintattico di essere esteso a nuovi elementi lessicali sulla base dell’uso che della costruzione è fatto dai parlanti.

Anche in questo caso, la produttività è analizzata come una proprietà scalare rappresentabile lungo un gradiente continuo ai cui estremi sono collocati rispettivamente la piena e la bassa produttività. Le categorie collocate nella parte più alta, oltre ad essere caratterizzate dalla massima produttività sono anche quelle più regolari e generali, ai piedi del gradiente si trovano invece processi produttivi di carattere analogico.

Secondo Barðdal, inoltre, le costruzioni si collocano lungo un continuum di schematicità-lessicalità in cui al livello più basso della gerarchia si trovano le istanze concrete, lessicalmente specificate e portatrici della maggior parte del contenuto semantico. Pian piano che i livelli salgono, si avrà un’astrazione sempre maggiore rispetto al livello precedente e una diminuzione dell’informazione semantica, fino a raggiungere il livello più alto della gerarchia in cui la costruzione è rappresentata in modo completamente schematico ed è portatrice di contenuto informativo di carattere esclusivamente relazionale.

Barðdal (2008), in opposizione con i modelli cognitivo-funzionali che parlavano della schematicità come nozione puramente linguistica ritiene che sia un concetto psicolinguistico fondamentale per determinare a quale livello, una determinata costruzione sia rappresentata nella mente del parlante. Il livello più alto di schematicità determina il più alto livello di produttività della costruzione.

The difference between highly productive and semi-productive patterns is not only a difference in the entrenchment of a super-ordinate schema, but rather a difference in both the type frequency of the two schemas and in the degree of schematicity at which they exist. (Barðdal, 2008:49)

Altri studi partono dagli assunti appena delineati e riconoscono nella produttività un fenomeno scalare e la proprietà delle costruzioni sintattiche di essere estese a nuove elementi lessicali (Israel, 1996; Bybee e Eddington, 2006; Matthwes e Bannard, 2009; Suttle e Goldberg, 2011; Perek, 2015, 2016; Theakston et al., 2014.).

2.3 Fattori di produttività sintattica nelle costruzioni a struttura argomentale.

Le varie teorie che si sono occupate del fenomeno della produttività, hanno presentato prospettive differenti riguardo a cosa si intenda con tale nozione, a cosa si applichi nello specifico e soprattutto a come possa essere quantificata e misurata. Dedicheremo questo paragrafo a delineare quali sono i fattori ritenuti responsabili del fenomeno della produttività e quali invece i vincoli che lo frenano, il modo in cui interagiscono fra di loro e la rispettiva importanza che i diversi studiosi conferiscono a ciascuno di essi.

2.3.1 Produttività, semantica lessicale e conoscenza del mondo.

Un aspetto fondamentale nella definizione della produttività sintattica risiede nell'importanza attribuita alla semantica lessicale e alla *world knowledge*.

Il fenomeno della selezione argomentale è stato da sempre considerato una proprietà intrinseca delle parole o dei morfemi. Pertanto, l'aspetto semantico ha rivestito un ruolo cruciale ed è stato analizzato dalla semantica lessicale, un ramo della linguistica che si occupa di studiare il significato delle parole sia da un punto di vista paradigmatico (le relazioni che intercorrono fra le parole *in absentia*) che da quello sintagmatico (relazioni fra parole *in praesentia*).

Nel processo di selezione argomentale le teste sintattiche da un lato specificano, a livello lessicale, la struttura argomentale semantica e come essa debba essere realizzata

a livello sintattico, dall'altro, impongono restrizioni semantiche sul paradigma di lessemi che possono riempire un determinato slot al fine di garantire un output semanticamente corretto.

Si veda l'esempio:

13) Maria cucina l'arrosto.

Il verbo *cucinare* richiede che determinati ruoli tematici siano istanziati da specifici argomenti sintattici, soddisfacendo il principio, proprio della grammatica generativa, del *theta-criterion* per cui ad ogni ruolo tematico deve corrispondere un argomento e viceversa.

Cucinare dunque apparirà in una struttura argomentale con due argomenti, un AGENTE e un TEMA: *cucinare*(AGENTE, TEMA). La stessa testa può apparire con diverse strutture argomentali che fanno riferimento ai medesimi ruoli tematici realizzati in modo differente a livello sintattico.

14) L'arrosto è stato cucinato da Maria.

Allo stesso tempo però, è necessario imporre determinati vincoli semantici al range di lessemi che possono essere istanza di un determinato slot sintattico.

15) ?? Maria cucina una pietra

Nonostante la struttura argomentale di (15) sia paragonabile a quella di (13) e la frase sia accettabile da un punto di vista grammaticale, quest'ultima, se non collocata all'interno di un determinato contesto pragmatico, risulterà sicuramente strana alle orecchie di qualsiasi parlante.

Numerosi sono gli approcci che hanno cercato di spiegare in che modo determinati verbi impongano restrizioni semantiche sul tipo di argomenti con i quali ricorrono (Katz e Fodor 1963, Dowty 1991, Jackendoff 1990, Wierzbicka 1996). Comune a tutti gli approcci è l'idea di postulare un insieme di vincoli semantici che operano sugli slot, siano essi definiti dallo slot stesso o una conseguenza delle relazioni sintagmatiche della costruzione. Tali vincoli sono strettamente correlati con la conoscenza che il parlante ha

del mondo che lo circonda, a tal punto da rendere talvolta difficoltosa la distinzione fra restrizioni di carattere concettuale-semantiche e vincoli di tipo pragmatico.

Quando utilizziamo un verbo come *bere*, la scelta dell'argomento che riveste il ruolo di PAZIENTE, dipende sia dalle nostre esigenze comunicative che dalla nostra conoscenza del mondo circa le cose che posso essere bevute e non; i parlanti sanno, grazie all'esperienza che hanno col mondo, che le sostanze che si bevono, faranno parte della classe semantico-concettuale [+liquid]. Tale conoscenza è utilizzata dai parlanti quando estendono il verbo a nuovi oggetti; la mancata appartenenza di un oggetto a una determinata classe spiega la malformazione di (16):

16) ?? Luca ha bevuto un treno_[-liquid]

Un aspetto altrettanto importante è rivestito dalle convenzioni con cui ogni singola lingua decide di descrivere a livello linguistico un determinato tipo di evento.

In italiano, siamo soliti dire *mangiare la zuppa*, mentre una lingua come il giapponese, per indicare la stessa azione, utilizza l'espressione *suupu o nonda*, equivalente a “*io bevo la zuppa*”, e lo stesso verbo è utilizzato per indicare l'ingestione di medicine, sia liquide che in pillole. In questo caso, il ricorso alla classe semantica [+liquid], non è sufficiente a spiegare la discrepanza di comportamenti fra le due lingue, dovuta invece a differenze di lessicalizzazione specifiche di ciascuna lingua.

Il rispetto dei soli vincoli semantici e pragmatici tuttavia non è sufficiente a spiegare interamente il fenomeno della realizzazione argomentale, e in particolare la predisposizione di determinati slot ad essere riempiti da un numero ed una varietà maggiore di argomenti rispetto ad altri.

L'importanza di tale componente idiosincratICA è stata recentemente riconosciuta dai modelli usage-based del linguaggio; essi ritengono, contrariamente a quanto sostenuto dai modelli generativi di stampo semanticista, che la produttività sintattica di una costruzione non sia totalmente predicibile a partire dalle proprietà semantico-lessicali delle sue componenti o dalla conoscenza che i parlanti hanno del mondo, e che determinate strutture argomentali siano più predisposte ad accettare nuovi elementi lessicali, indipendentemente dalla loro appartenenza ad una determinata classe semantica concettuale.

Per dimostrare tale componente idiosincratICA, Zeldes (2012) ha analizzato le differenze di produttività sia di gruppi di verbi inglesi sinonimici sia di costruzioni sinonimiche al

cui interno ricorrono verbi o teste sintattiche le cui differenze di tipo pragmatico o di registro sono ridotte al minimo. Un caso interessante è rappresentato dalla costruzione tedesca costituita da *wegen* (a causa di) in posizione preposizionale o posposizionale, solitamente accompagnata dal genitivo, o, nel caso della prima variante, anche dal dativo (in contesti colloquiali). Le due costruzioni non differiscono per quanto riguarda l'aspetto semantico, entrambe sono attestate nel tedesco contemporaneo ma differiscono nel grado di formalità e colloquialismo: la variante con *wegen* posposizionale è più arcaica e costituisce un'espressione più fissa e formulaica mentre la seconda variante è più recente e aperta ad accettare nuovi elementi lessicali. L'idea alla base dello studio di Zeldes è che, se la semantica lessicale e la world knowledge fossero sufficienti a spiegare la selezione di nuovi argomenti, costruzioni che hanno lo stesso significato e lo stesso scopo pragmatico dovrebbero mostrare pattern di produttività simili. I risultati della sua analisi, mostrano al contrario che le due costruzioni, pur essendo entrambe produttive, lo sono in gradi differenti.

Zeldes, si ricollega agli studi compiuti da Bayeen sulla produttività dei processi morfologici, riprendendo quattro misure: la frequenza token ($N(C)$), la frequenza type (V), \mathcal{P} e S . Come abbiamo visto nel paragrafo precedente, \mathcal{P} è un indice di produttività e indica il rapporto fra il numero di hapax legomena e il numero di tokens, mentre S indica la misura del limite asintotico di crescita di V di un dato processo, ovvero, il valore massimo di V che ci aspetteremmo di trovare se continuassimo a leggere un testo dello stesso genere per un tempo illimitato.

Se la prima e l'ultima sono proprietà invariabili per ciascuno slot, V e \mathcal{P} dipendono invece dalla dimensione del campione di testo preso in analisi. Dal momento che la dimensione del vocabolario di un processo può essere particolarmente ampia, ma i singoli type relativamente rari, Zeldes utilizza i modelli statistici LNRE (*large number of rare events*) per misurare le proprietà che dipendono dalla crescita del vocabolario.

Attraverso tali misurazioni, egli dimostra che la variante posposizionale è meno produttiva, presenta un vocabolario più ridotto e ripetitivo, e, in proporzione, un numero di hapax legomena su un campione di testo equivalente, più basso.

È importante sottolineare come tali risultati non siano volti a delegittimare l'importanza dell'aspetto semantico nel fenomeno della realizzazione argomentale. L'intento dello studioso è, al contrario, quello di rifiutare una visione assolutistica per cui quest'ultima dipenderebbe esclusivamente da vincoli imposti dalle classi semantiche; sono numerosi i fattori che contribuiscono a rendere produttiva una costruzione e questi riguardano

l'aspetto distribuzionale e la frequenza con cui una determinata espressione e i suoi elementi costitutivi ricorrono nella lingua.

Lo scopo dei modelli usage-based, oltre a quello di fornire una definizione del concetto di produttività, è offrire una serie di criteri qualitativi e quantitativi per misurarlo e per farlo spesso ricorrono a nozioni di carattere distribuzionale.

Questo secondo tipo di informazioni sono memorizzate nella mente del parlante e lo guidano durante il processo di produzione nel selezionare la costruzione più adatta al contesto linguistico e comunicativo sulla base degli usi che della costruzione sono stati fatti precedentemente.

2.3.2 Misure di produttività sintattica.

Abbiamo visto, nei paragrafi precedenti come la produttività sia sottoposta a vincoli che regolano la possibilità che una parola compaia all'interno di una determinata costruzione. Tale fenomeno, in letteratura, è insito nella nozione di *produttività parziale* (Goldberg, 1995) e in quella di estendibilità di cui parla Barðdal (2008).

Secondo Suttle e Goldberg (2011) sono due i criteri minimi che devono essere rispettati affinché un nuovo conio possa entrare a fare parte di una lingua.

Prima di tutto deve rispondere al criterio della *semantic sensicality*, cioè deve avere un significato a livello semantico. Ciò è facilmente comprensibile anche a livello intuitivo: i parlanti non producono espressioni prive di senso perché altrimenti verrebbero meno allo scopo principale della comunicazione umana, essere compresi dagli altri.

Affinché tale criterio sia rispettato, Goldberg (1995) postula l'esistenza di vincoli che regolano la combinazione fra i verbi e le costruzioni (*Semantic Coherence Principle* e *Correspondence Principle*, cfr. Goldberg, 1995), e sottolinea tuttavia l'importanza rivestita dal contesto, che talvolta interviene per fornire contributo informativo e per migliorare formazioni che altrimenti risulterebbero malformate.

L'altro criterio è rappresentato dalla *pre-emption* statistica (Pinker, 1981; Goldberg, 1995, 2006; Brooks e Tomasello, 1999; Boyd e Goldberg, 2010; Suttle e Goldberg, 2011): un nuovo conio, non deve essere bloccato da altre costruzioni già presenti nella lingua che svolgano la medesima funzione o una funzione più appropriata.

La pre-emption in un certo senso rappresenta una forma di evidenza negativa indiretta e si basa sull'idea che una conoscenza più specifica blocchi una conoscenza più generale

purché entrambe soddisfino i bisogni funzionali del contesto e condividano gli stessi vincoli semantici e pragmatici.

Pre-emption is a particular type of indirect negative evidence that results from repeatedly hearing a formulation, B, in a context where one might have expected to hear a semantically and pragmatically related formulation, A. Given this type of input, speakers implicitly recognize that B is the appropriate formulation in such a context, and that A is not appropriate. (Boyd e Goldberg, 2010: 12)

Se l'effetto della pre-emption è stato analizzato e testato empiricamente in campo morfologico, per quanto riguarda i pattern sintattici, tale fenomeno è più difficile da dimostrare per il fatto che è difficile trovare casi di piena sinonimia semantica e pragmatica; spesso, ciò che distingue due costruzioni sintattiche, è il diverso modo in cui è "impacchettata" la struttura informativa.

Si prenda il caso della costruzione ditransitiva inglese e della sua variante preposizionale:

- 17) a. ?? She explained someone the story.
b. She explained the story to someone.

- 18) a. She told someone the story.
b. She told the story to someone.

Le due costruzioni presentano un modo differente di impacchettare la struttura informativa, inoltre la maggior parte dei verbi può occorrere indistintamente nell'una e nell'altra costruzione. Cosa blocca, allora, in determinati casi, i parlanti nell'utilizzare un pattern piuttosto che un altro?

Goldberg ricorre alla nozione di pre-emption statistica:

Learners may witness repeated situations in which the ditransitive might be expected because the relevant information structure suits the ditransitive at least as well as the prepositional paraphrase. If, in these situations, the prepositional alternative is systematically witnessed instead, the learner can implicitly infer that the ditransitive is not after all appropriate. (Boyd e Goldberg, 2010: 13)

Tale fenomeno è essenzialmente statistico, affinché possa realizzarsi, un ruolo fondamentale è svolto dalla frequenza con cui il parlante è esposto ad una determinata costruzione linguistica: solo dopo la continua esposizione ad una costruzione piuttosto che ad un'altra il parlante comprende che la seconda non è accettabile nella lingua.

I due criteri appena analizzati, la *semantic sensicality* e la *preeption statistica*, interagiscono dunque per vincolare gli usi creativi delle parole all'interno delle costruzioni ma da soli non sono sufficienti; esistono costruzioni che hanno una frequenza troppo bassa o sono troppo specializzate a livello semantico e pragmatico perché si verifichi il fenomeno della pre-emption e possano essere quindi bloccate da un'altra costruzione.

Sono stati individuati altri quattro fattori che, in varia misura, forniscono contributi differenti e graduati alla produttività delle costruzioni:

- a. *frequenza token*
- b. *frequenza type*
- c. *variabilità semantica*
- d. *similarità*

Frequenza Token. La frequenza token rappresenta il numero di volte con cui un elemento compare all'interno di uno slot o alla frequenza di un pattern all'interno di un testo o di un corpus. La struttura di una costruzione è abbastanza complessa: abbiamo visto come essa sia costituita da elementi che rimangono costanti e invariabili e da posizioni aperte (slot) che possono essere riempite da item lessicali differenti nel rispetto dei vincoli semantici imposti dalla costruzione.

Ogni volta che una costruzione è utilizzata, le parti che al suo interno rimangono costanti sono rafforzate (*entrenched*) nella rappresentazione mentale che il parlante ha di tale costruzione e si rendono disponibili come modelli analogici per nuove espressioni.

Tale fattore ha spesso costituito per gli studiosi un vincolo al fenomeno della produttività e sono numerosi gli studi che hanno cercato di dimostrare il ruolo della frequenza token nel vincolare la generalizzazione di una costruzione (Theakston, 2004; Ambridge et al., 2008; Ambridge et al., 2011). Secondo Bybee (2010), elementi contraddistinti da una frequenza token molto alta possono formare un *chunk* autonomo e non attivare più l'intero *exemplar cluster* della costruzione corrispondente.

[...] the more often a sequence is processed directly as a unit, the less likely it is to activate other units or the construction to which it belongs and the more likely it is to lose its analysability. At the same time, use in particular contexts contributes to shifts in meaning, which decrease compositionality and make the former exemplar of a construction move away from its source. (Bybee, 2010:96)

Un punto di vista alternativo circa il ruolo svolto dalla frequenza token nell'estensione delle costruzioni, è dovuto a Barðdal (2008), la quale ritiene che quest'ultima sia un fattore psicolinguistico di fondamentale importanza nell'individuare gli elementi a partire dai quali i parlanti compiono estensioni analogiche.

Each token also instantiates a type, and if token similarity contributes to analogical extensions, then the token frequency of the possible model verbs should be of utmost importance. [...]

Type frequency will be an indicator of the highest level of schematicity each construction exists at, and hence an indicator of the semantic scope of the construction and its productivity domain, but token frequency will be an important psycholinguistic factor singling out model items for speakers when they extend low-level constructions. (Barðdal, 2008:93-96)

Se da una parte, costituisce dunque un elemento inibitorio per la produttività delle costruzioni più generali, dall'altra costituisce un fattore determinante per le estensioni di carattere analogico.

Secondo Barðdal, fra frequenza token e frequenza type si delinea così una relazione di correlazione inversa.

Frequenza Type. Già individuata come fattore importante nella determinazione della produttività morfologica (Bybee, 1985), la frequenza type rappresenta un elemento centrale nella definizione della produttività sintattica e costituisce una proprietà delle costruzioni formate da posizioni aperte (slot). Secondo una prospettiva usage-based, per cui la rappresentazione cognitiva è influenzata dalla frequenza con cui vengono utilizzate le singole unità della lingua, essa si riferisce al numero dei diversi elementi lessicali che possono comparire all'interno delle posizioni aperte delle costruzioni. Numerosi sono gli studiosi che hanno ritenuto che la frequenza type di una particolare costruzione, giochi un ruolo cruciale nel determinare la probabilità che tale processo

venga esteso a nuove forme (Bybee, 1995; Bybee e Thompson, 1997; Goldberg, 2006; Barðdal, 2008; Suttle e Goldberg, 2011; Zeschel, 2012; Perek, 2015, 2016)

Maggiore è la frequenza type di una costruzione, più probabilità c'è che tale costruzione sia estesa a nuovi elementi lessicali:

Constructions that have been heard used with a wide variety of types are more likely to be extended broadly than constructions that have been heard used with a semantically circumscribed set of types. (Goldberg, 2006: 99)

La frequenza type costituisce un importante strumento che i bambini e gli apprendenti adulti utilizzano nel processo di acquisizione del linguaggio ed è utilizzata costantemente dai parlanti adulti per quantificare il grado di accettabilità dell'estensione di un pattern sintattico ad un nuovo elemento lessicale.

Variabilità semantica e coerenza semantica. Il grado di variabilità corrisponde alla gamma delle istanze attestate in una costruzione. Tale fattore costituisce un indice dell'apertura della costruzione e della sua predisposizione ad accettare nuovo materiale lessicale al suo interno: le costruzioni che sono utilizzate con una grande varietà di elementi lessicali diversi, sono più propense alle estensioni di quelle circoscritte ad un gruppo particolare di types.

Il concetto di variabilità è collegato a quello di coerenza semantica che Barðdal (2008) definisce *coerenza* semantica di uno schema (“the internal consistency found between all the members of each schema or a category”, 2008:27), entrambi infatti analizzano lo stesso fenomeno da prospettive differenti: maggiore è la variabilità delle istanze minore sarà il grado di coerenza semantica di un pattern.

Similarità semantica con istanze già attestate e la nozione di coverage. Un fattore molto importante che contribuisce a determinare la produttività di una costruzione è costituito dal grado di similarità semantica fra la nuova istanza e le istanze già attestate (Bybee e Eddington, 2006; Goldberg, 2006; Barðdal, 2008; Matthews e Bannard, 2009; Suttle e Goldberg, 2011; Perek, 2015, 2016).

Learners are fairly cautious in producing utterances based on generalizing beyond the input. They can only be expected confidently to use a new verb in a familiar pattern when that new verb is relevantly close in meaning to verbs they have already heard used in the pattern. (Goldberg, 2006:99)

I parlanti sono più predisposti ad accettare una nuova istanza all'interno di una determinata costruzione nella misura in cui essa è semanticamente simile alle istanze già attestate.

Il rapporto fra la variabilità e la similarità semantica può essere espresso attraverso il concetto di *coverage*, già utilizzata da Osherson et al. (1990) per parlare dell'induzione categoriale.

Osherson et al. hanno dimostrato che le persone sono più portate a generalizzare una proprietà ad una nuova istanza nella misura in cui l'istanza rientra nella categoria costituita da elementi che già esibiscono tale proprietà.

Nel caso del linguaggio, l'analogia è chiara:

The greater the degree to which previously attested instances fill a semantic space that include the potential target instance, the more confident speakers will be in using the target instance.

(Goldberg 2006: 100)

La *coverage* è definita da Goldberg (2006) e Suttle e Goldberg (2011) come un fattore scalare e indica il grado in cui le istanze attestate ricoprono una certa categoria determinata unitamente dalle istanze attestate e dal target della nuova estensione.

Tale nozione è collegata alla densità del dominio semantico di una costruzione, cioè alla misura in cui il dominio semantico di una costruzione è popolato nelle vicinanze dell'istanza target; se lo spazio semantico è denso, ovvero è popolato da un numero elevato di elementi lessicali semanticamente simili, la probabilità che venga accolto dalla costruzione è elevato; al contrario, più è rado lo spazio semantico, minore sarà la probabilità che l'istanza compaia all'interno della costruzione.

I vari studi che sono stati compiuti in questo ambito si differenziano per quanto riguarda la relativa importanza che ognuno di essi attribuisce a ciascun fattore e nel modo in cui questi ultimi interagiscono gli uni con gli altri.

2.3.3 Studi precedenti.

Per dimostrare l'effetto della pre-emption statistica in sintassi, Boyd e Goldberg (2010) hanno compiuto uno studio sull'utilizzo degli aggettivi inglesi che iniziano per schwa, ə, (*afraid*). In inglese, tali aggettivi, non compaiono in strutture attributive in posizione prenominali (*the scared boy* vs. ??*the afraid boy*) senza però che tale

restrizione possa essere attribuita a fattori di carattere semantico o fonologico. Lo scopo dello studio era dunque quello di analizzare cosa blocca i parlanti nell'utilizzare determinate strutture linguistiche che sarebbero, in linea teorica, permesse dalla lingua e accettabili da un punto di vista semantico.

Boyd e Goldberg hanno dunque esposto i soggetti all'utilizzo di due coni aggettivali all'interno di una struttura relativa (19), e hanno indagato se tale struttura potesse rivestire un ruolo preventivo, bloccando l'utilizzo di tali aggettivi all'interno della struttura attributiva (20).

19) The pig that was *ablim* moved to the star.

20) The *ablim* pig moved to the star.

L'ipotesi dei due studiosi era che l'esposizione alla struttura relativa (19) *nell'exposure block*, fornisse ai partecipanti un'evidenza negativa indiretta, bloccandoli nell'utilizzo della struttura attributiva nel task di produzione.

I risultati ottenuti, oltre a dimostrare la validità di tale ipotesi, mostrano come la pre-emption statistica giochi un ruolo cruciale nel processo di acquisizione linguistica e di generalizzazione, portando i parlanti a generalizzare le restrizioni a tutti i membri della categoria e non solo a quelli osservati direttamente.

Bybee e Eddington (2006) hanno analizzato il ruolo svolto dalla frequenza token nel guidare i parlanti ad esprimere giudizi di accettabilità di coppie di verbi di cambiamento di stato seguiti da aggettivi (nella fattispecie i verbi analizzati sono *quedarse*, *ponerse*, *volverse*, *hacerse*). Attraverso una prima analisi, gli studiosi hanno evidenziato come nei dati estratti dai corpora, determinate coppie, caratterizzate da una frequenza elevata, sembrano costituire "prefabs" o espressioni fisse dal carattere ormai convenzionalizzato e, allo stesso tempo, come gli altri aggettivi che ricorrono nel pattern con frequenza più bassa, siano per la maggior parte sinonimi delle espressioni utilizzate in tali espressioni convenzionalizzate o condividano con esse numerose caratteristiche semantiche.

L'ipotesi postulata dai due studiosi è che gli elementi che ricorrono con una frequenza token elevata all'interno della costruzione, fungano da membri centrali, prototipici delle categorie che sono istanza dello slot schematico della costruzione stessa e costituiscano gli elementi a partire dai quali avviene il processo di estensione.

Numerosi sono gli studi che hanno analizzato il ruolo svolto dalla frequenza type nel fenomeno della produttività.

Zeschel (2010) ha proposto uno studio corpus-based volto ad indagare gli effetti della frequenza type nel processo di estensione semantica di collocazioni del tipo Agg. + N, caratterizzate dalla presenza dell'aggettivo tedesco *tief*, particolarmente polisemico. L'ipotesi di partenza era che la distribuzione di nuove combinazioni Agg. + N nello spazio semantico fosse correlata positivamente con la distribuzione delle combinazioni già attestate.

I dati sono stati sottoposti ad una analisi statistica per verificare la correlazione fra la distribuzione dei pattern attestati e di quelli nuovi per ciascuna variante semantica. I risultati hanno mostrato una correlazione positiva fra la *semantic type frequency* delle varianti attestate di una collocazione e il numero delle estensioni creative, dimostrando una effettiva connessione fra la frequenza type di una costruzione e la probabilità che i parlanti la estendano a nuove istanze lessicali.

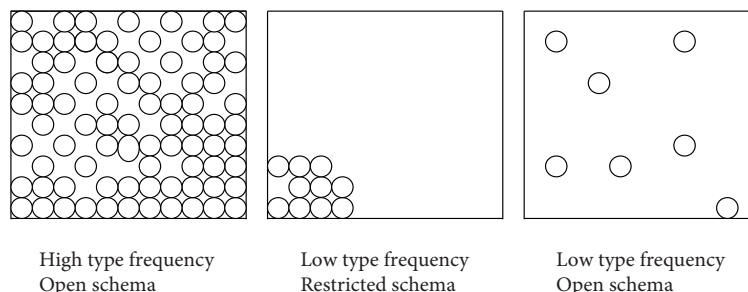
Barðdal (2006, 2008) in uno studio sulle costruzioni a struttura argomentale islandesi ha analizzato la correlazione fra frequenza type e coerenza semantica (ovvero la coesione interna tra gli elementi che compongono la classe di riferimento per uno slot di una costruzione) individuando una relazione inversa fra i due fattori nel determinarne il grado di produttività. Maggiore è la frequenza type di una costruzione, minore sarà il grado di coerenza semantica degli elementi che sono istanza degli slot aperti e viceversa.

For a category that is high in type frequency, this means that only a low degree of internal consistency across its items is needed for the category to be extended. Categories of intermediate type frequency have to be more coherent than high type-frequency categories in order to be extended to new items. Finally, categories that are low in type frequency must show the highest degree of internal consistency across their items in order to be extended to new types. (Barðdal, 2006:469)

Per testare la propria ipotesi, la studiosa ha analizzato la distribuzione di prestiti verbali presi dall'area dell'*information technology* all'interno di due costruzioni a struttura argomentale islandesi: Nom.-Acc. e Nom-Dat. Barðdal ha individuato, in questo modo, tre punti del *productivity cline* che rappresentano tre livelli emblematici di correlazione e sono rappresentati rispettivamente dalla costruzione Nom.-Acc., semanticamente aperta e caratterizzata da un'elevata frequenza type, dalla costruzione Nom.-Dat., più circoscritta a livello semantico e con livelli di frequenza type intermedi ed infine il caso isolato di un conio verbale (*dona uppi*) accompagnato da una particella preposizionale

e caratterizzato da un significato altamente specifico, che secondo la studiosa si è formato per analogia con una forma già presente nella lingua (*daga uppi*).

Figura 7 Correlazione inversa fra frequenza type e coerenza semantica secondo Barðal (2008: 43).



Recentemente, Perek (2015) ha indagato il ruolo svolto dalla frequenza type nel determinare la produttività nel fenomeno delle alternanze argomentali. Partendo dalla constatazione che esistono asimmetrie nella produttività delle varianti di un’alternanza argomentale, lo studioso si è domandato se l’esperienza che i parlanti hanno con un determinato verbo all’interno di una delle varianti dell’alternanza influisca sulla probabilità che lo stesso verbo compaia anche nell’altra. Lo studio è stato condotto utilizzando come stimoli le varianti dell’alternanza locativa e di quella dativa. Ai soggetti sono stati presentati coni verbali in una delle due varianti di ciascuna alternanza, ed è stato chiesto loro, successivamente, di utilizzarli all’interno di una frase (*sentence completion task*); lo studioso ha poi condotto una misurazione della frequenza con cui tali verbi sono stati utilizzati dai parlanti nell’altra variante.

I risultati dell’esperimento mostrano che per quanto riguarda l’alternanza dativa, i parlanti mantengono un comportamento conservatore se il verbo è incontrato, la prima volta, nella variante “to-dativo” ma non nel caso in cui esso sia presentato per la prima volta nella variante ditransitiva; in questa condizione, i parlanti hanno, al contrario, un comportamento produttivo e sembrano pronti ad estendere il verbo ad entrambe le varianti indipendentemente da quella esperita nello stimolo. Non sembrano esserci invece forme di asimmetria per quanto riguarda l’alternanza locativa. La spiegazione fornita dallo studioso, in linea con una prospettiva usage-based, attribuisce la differenza di produttività ad una differenza di frequenza type non solo delle singole varianti ma anche dell’intero *constructeme*.

The existence of many more verbs occurring only in the to-dative construction than occurring in both variants of the alternation triggers the expectation that a verb used in the

to-dative cannot be used in the ditransitive; conversely, the low number of ditransitive-only verbs triggers the expectation that a verb used in the ditransitive can also be used in the to-dative. (Perek, 2015: 198)

Per quanto riguarda l'alternanza locativa, il processo è analogo: i soggetti hanno un comportamento conservatore in entrambe le condizioni poiché il numero dei verbi che ricorre in una delle due varianti è maggiore rispetto a quelle dei verbi che ricorrono in entrambe.

Un aspetto importante sottolineato da Perek è che la produttività non è dunque un fenomeno che interessa le singole costruzioni, ma ha a che fare con generalizzazioni cross-costruzionali e può essere spiegata se le varianti delle alternanze argomentali sono trattate in termini di allostructions; in questo modo, la produttività non dipende dalla frequenza type di una singola variante ma da quella della constructeme (cfr. Cappelle, 2006), intesa come unità di conoscenza linguistica.

Un'ulteriore spiegazione fornita da Perek circa l'effettiva asimmetria nella produttività delle alternanze argomentali potrebbe essere legata al fenomeno della pre-emption statistica: secondo lo studioso, i soggetti sarebbero restii ad estendere produttivamente i verbi incontrati primariamente nella variante "to-dativo" alla variante ditransitiva, poiché sarebbero vincolati all'esperienza che del verbo è stata fatta in precedenza. Resta tuttavia da analizzare che tipo di relazione intercorre fra i due fenomeni, produttività e pre-emption.

Risultati analoghi sono stati riscontrati negli studi di Wannacott e al. (2008), Wannacott (2011) e Perek e Goldberg (2015).

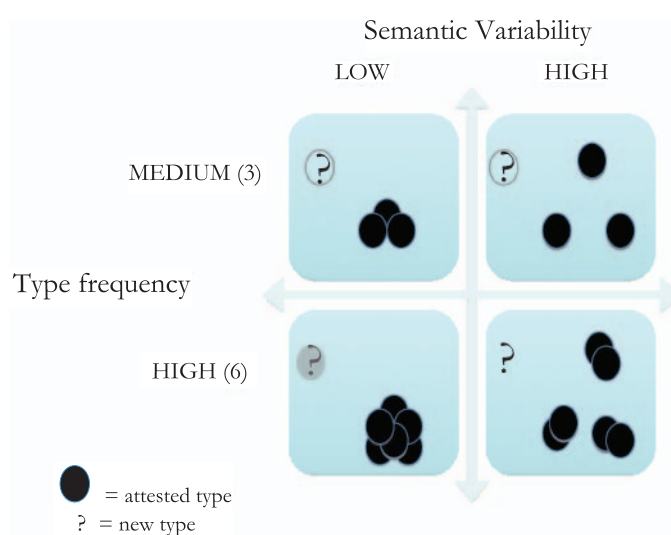
Uno studio particolarmente interessante è stato compiuto recentemente da Suttle e Goldberg (2011) le quali hanno analizzato il contributo di frequenza types, variabilità e similarità semantica nel determinare la produttività sintattica e il modo in cui questi tre fattori interagiscono fra loro.

Attraverso tre esperimenti condotti utilizzando una lingua inventata, lo "Zargotian", le due studiose hanno cercato di individuare quale fosse il ruolo della frequenza type e della variabilità semantica nel determinare la produttività di una costruzione grammaticale e il grado d'interazione fra i due fattori. Ai soggetti sono stati presentati gruppi di frasi costituite da due sintagmi nominali, formati da un articolo determinativo e un nome nonsense in lingua "zargoziana", seguite da un verbo inglese accompagnato da una particella anch'essa nonsense: *The zask the nop toast-pe*. Per ogni set, bassa, media e alta frequenza type sono state incrociate a livelli di bassa e alta variabilità. I

soggetti sono stati esposti ad un set di frasi dopodiché è stato chiesto loro di formulare un giudizio di accettabilità riguardo ad una frase target. Il primo fattore è stato controllato variando il numero dei verbi all'interno di ciascun set, mentre la variabilità è stata manipolata scegliendo gli stimoli verbali da classi semantiche diverse in accordo con la classificazione di Levin (1993), la cui similarità è stata controllata utilizzando la Latent Semantic Analysis .

I risultati degli esperimenti hanno evidenziato un effetto positivo per l'elevata variabilità e l'elevata frequenza type, che sembrano avere effetti indipendenti nell'incoraggiare la produttività delle costruzioni e interagire positivamente fra di loro.

Figura 8 Rapporto fra variabilità semantica e frequenza type in Goldberg et al. (2011:1245).



Il concetto di variabilità semantica di cui parlano Suttle e Goldberg può essere collegato a quella che Barðdal (2008) definisce *coerenza* semantica di uno schema: entrambi, come abbiamo visto analizzano lo stesso fenomeno da prospettive differenti. Tuttavia, il distacco fra Suttle e Goldberg (2011) e Barðdal (2008) non è solo di natura terminologica o prospettica. Le prime infatti, sottolineano l'importanza di distinguere fra la coerenza semantica di una costruzione e quella degli elementi lessicali (verbi) che ne possono essere istanza: una costruzione può infatti presentare una elevata coerenza semantica ma accogliere al suo interno una gamma variabile di verbi (cfr. la *way construction*, analizzata da Israel, 1996). Inoltre, se Barðdal sottolineava l'importanza della correlazione fra frequenza type e coerenza semantica nella predizione della produttività, Suttle e Goldberg (2011) pur riconoscendone l'interazione nelle lingue naturali, analizzano il contributo dei due fattori separatamente.

Le due studiose hanno poi compiuto un terzo esperimento volto a indagare come i tre fattori interagiscano fra loro (frequenza type, variabilità e similarità).

Nell'esperimento, sono state create tre condizioni incrociando tre livelli di similarità (alta, media e bassa) del conio con le istanze attestate, a due livelli di variabilità (alta e bassa). I verbi utilizzati per creare gli stimoli sono stati scelti, dalle varie classi semantiche utilizzate nell'esperimento precedente. I risultati dell'esperimento hanno evidenziato un'interazione complessa fra variabilità e similarità semantica.

In particolare hanno riscontrato che l'effetto della variabilità varia in risposta al variare del grado similarità della nuova istanza con istanze già attestate. Con un livello di similarità medio, si riscontrava un effetto positivo della variabilità, ciò significa che quando la nuova istanza proviene da una classe verbale differente, ma non totalmente scollegata a livello semantico, rispetto alle istanze attestate, i soggetti sono più portati ad estendere la costruzione ad un nuovo conio se quest'ultima è già attestata con istanze semanticamente variabili. Nella condizione di elevata similarità, l'estensione del pattern sintattico a nuove istanze è stato giudicato meno accettabile dai parlanti, quando correlato ad un'elevata variabilità delle istanze attestate; in questa condizione i soggetti, sembrano più predisposti a estendere un pattern ad un nuovo elemento lessicale, se la distribuzione della costruzione è meno variabile. Quando infine il parametro della similarità è basso, non è stato riscontrato alcun effetto della variabilità, sottolineando come quest'ultima sia irrilevante quando la nuova istanza non ha alcuna somiglianza con istanze attestate.

Per spiegare questa interazione, le due studiose ricorrono al concetto di *coverage*, che indica il grado in cui le istanze attestate ricoprono una certa categoria determinata unitamente dalle istanze attestate e dal target della nuova estensione.

Ricollegando tale nozione al rapporto fra variabilità e similarità, il caso in cui la variabilità riduce l'effetto dell'elevata similarità semantica è spiegato in questo modo:

In that case, variability served to extend the required category well beyond what would otherwise be needed (if all instances and coinage came from the same semantic class). In other words, if the coinage is from the same category as all of the attested instances (high similarity), coverage is already maximized; higher variability can only serve to spread the category out, reducing coverage. [...] In the medium similarity condition, and only in this condition, the category created by combination of attested instances and target coinage covers the semantic space better when the attested instances are varied. (Suttle e Goldberg, 2011:1256)

Da questo scaturisce che la produttività sintattica non è una proprietà assoluta di una costruzione ma un fenomeno che tiene conto della relazione fra le istanze attestate e quelle potenziali. Secondo la nozione di coverage, infatti, una costruzione non può essere detta produttiva in termini assoluti ma lo è in gradi differenti e in domini semantici differenti.

Bybee e Eddington (2006) all'interno del medesimo studio volto ad indagare la distribuzione dei quattro verbi di cambiamento di stato in spagnolo (*quedarse*, *ponerse*, *hacerse* e *volverse*) hanno condotto un secondo esperimento per analizzare il relativo contributo della frequenza token e della similarità alle istanze attestate, nell'esprimere giudizi di accettabilità. Delle quattro combinazioni Verbo + Agg., sono state selezionate quelle al cui interno ricorrevano i verbi *quedarse* e *ponerse*, caratterizzati da una frequenza token e frequenza type più elevate. Ai partecipanti, madrelingua spagnoli, sono state presentate delle frasi al cui interno occorre le combinazioni target, riguardo alle quali è stato chiesto loro di fornire dei giudizi di accettabilità sulla base di una scala Likert. Per focalizzare l'attenzione dei soggetti sulla parte della frase d'interesse, e per evitare ogni tipo di influenza da parte di altri elementi, quest'ultima è stata sottolineata: es. *Con el vino y la voz pegajosa se estaba poniendo sentimental*. I risultati hanno mostrato che i giudizi di accettabilità sono influenzati sia dalla frequenza di ricorrenza delle combinazioni nella lingua, sia dalla similarità delle istanze meno attestate con le istanze più frequenti.

The high-frequency expressions and the low-frequency expressions that were semantically related to the high-frequency ones were judge more acceptable than the isolated (low frequency unrelated) combinations. This result indirectly confirms the centrality of the higher-frequency expressions in forming similarity clusters, as both the high-frequency and lower-frequency related expressions were judged to be more acceptable than the lower-frequency isolated expressions. (Bybee e Eddington, 2006:352)

Il ruolo della similarità nel determinare la produttività sintattica è stato riconosciuto recentemente anche da Perek (2016), il quale ha analizzato la produttività della costruzione “V *the hell out of* NP”, sulla base di dati semantico-distribuzionali. Tale costruzione è utilizzata solitamente per sottolineare l'effetto dell'azione espressa dal verbo o lo sforzo compiuto dall'agente nel metterla in atto e si costruisce con verbi appartenenti a due grossi domini semantici: verbi di cognizione e verbi di azione violenta.

Il pattern non può essere derivato composizionalmente da altre costruzioni, il significato “intensificatore” è associato direttamente all’intero pattern frasale ed esso dà origine alle proprie generalizzazioni.

Sulla base di analisi quantitative è emerso che la probabilità che un nuovo conio entri a far parte della costruzione dipende dalla similarità che quest’ultimo ha con le istanze già attestate al suo interno e dalla loro frequenza di ricorrenza, ovvero con la densità dello spazio semantico attorno a sé.

I risultati inoltre mostrano che una misura di densità supportata da un maggior numero di types è in grado di fare previsioni migliori rispetto ad una supportata da un numero di types più ridotto.

This means that productive use not only depend on whether the meaning of a potential coinage is sufficiently related to the existing semantic space, they also occur more reliably when this relation is supported by many items. These findings point to complementary roles that the semantic distribution of a construction and its type frequency play in syntactic productivity, as argued by Barðdal (2008): the former sets the necessary conditions for a new coinage to occur, while the type frequency increases the confidence that this coinage is indeed possible. (Perek, 2016: 25)

2.3.4 Una rappresentazione cognitiva della produttività sintattica: la legge di Hebb.

Numerosi sono gli studi che hanno cercato di fornire una spiegazione del modo in cui il fenomeno della produttività è rappresentato nella mente dei parlanti.

Uno dei contributi più interessanti risale a Zeldes (2012) il quale propone una visione della produttività come epifenomeno dell’organizzazione lessicale della lingua e di come essa è acquisita tramite l’esperienza. Le informazioni circa la produttività di una costruzione sono memorizzate nel lessico mentale insieme alle costruzioni stesse e ad informazioni circa il loro uso e la loro distribuzione.

Egli ricorre ad un noto modello di apprendimento che sta alla base di numerosi approcci connessionisti, la legge di Hebb (1949) la quale postula che se un neurone A attiva un neurone B ripetutamente, la connessione fra i due neuroni risulta rafforzata, al contrario, se due cellule non si attivano simultaneamente o si inibiscono rispettivamente, fra le due si instaurerà una relazione inibitoria. L’implicazione fondamentale che sta alla base di tale assunto è che l’uso modifica fisicamente l’architettura del cervello in quanto

rafforza quei circuiti neuronali utilizzati più frequentemente, predisponendoli anche ad utilizzi successivi.

Applicato alla rappresentazione delle costruzioni, la legge sembra suggerire che gli argomenti più frequenti, ed in particolare le collocazioni, sviluppino connessioni robuste con le loro teste, mentre gli argomenti meno frequenti, legami più deboli; questi ultimi proprio per il fatto di essere poco radicati andranno a rafforzare la rappresentazione della costruzione astratta. Allo stesso modo, gli slot argomentali che sono attestati con un numero elevato di argomenti rari (o *hapax legomena*), avranno maggiore facilità ad attrarre nuovi elementi poiché non sono sufficientemente robusti per essere associati ad una costruzione in modo esclusivo.

In questo modo la produttività non sarà altro che:

A function of how well the relevant construction is networked to conceptual structures, how strong its representation is, and at lower hierarchical levels, how well networked its lexicalized exemplars are. (Zeldes, 2012: 214)

Una frequenza token elevata dunque andrà a rafforzare il radicamento della costruzione nella mente del parlante mentre un'elevata frequenza type, corrisponderà alla presenza di un numero maggiore di *sub-network* in grado di attrarre nuovi elementi lessicali, infine una distribuzione asimmetrica, contribuirà da una parte a radicare la costruzione attorno a determinati esemplari che fungeranno da poli prototipici per eventuali estensioni di tipo analogico sulla base della similarità semantica, dall'altra, eviterà che tali esemplari prototipici vincolino la creazione di nuove forme.

The Hebbian account of productive construction acquisition from skewed input containing a high proportion of rare types as contrasted with less productive, though possibly frequent constructions, is both cognitively plausible and predictive of the multiple aspects of productivity we have seen. The resulting network of dynamically updated representations biases speakers during the selectional process away from certain less productive constructions when called upon to embed material compositionally, and towards others that are more productive. Conversely, collocational arguments will favour the constructions in which they are familiar, though of course a wide range of factors, from priming, through information structure to register, can also play a role in the competition of activation between the sub-network. (Zeldes, 2012: 218)

2.4 Sintesi.

Abbiamo visto come uno studio sulla produttività linguistica deve dunque analizzare non solo quali sono i principi alla base della creazione dei pattern linguistici ma anche in che modo i parlanti sono in grado di adattare questi ultimi a nuovi contesti di applicazione.

L'obiettivo delle teorie costruzioniste, di stampo usage-based, è dunque quello di indagare come nasce la conoscenza della produttività, come è acquisita, dove e come è rappresentata nella mente del parlante; se tale nozione sia parte della conoscenza grammaticale o invece il risultato dell'interazione fra componenti linguistiche, struttura semantica concettuale e conoscenza del mondo; e di indagare come mai certe costruzioni sono facilmente estendibili a nuovo materiale lessicale mentre altre mantengono una maggiore rigidità; perché varianti di un determinato tipo semantico sono accettabili in misure differenti all'interno di un medesimo pattern; e infine, quale sia il meccanismo alla base di tali estensioni e a quale livello di astrazione è codificata tale conoscenza.

Abbiamo sottolineato come spesso il concetto di produttività sia affiancato da quello di schematicità.

Con l'aumentare/diminuire della schematicità, il significato di una costruzione diventa rispettivamente più generale o più specifico. Un significato più schematico può applicarsi ad uno spettro più ampio di situazioni, ed un maggior numero di elementi possono diventare istanza dello schema stesso.

L'occorrenza di nuovi types all'interno della costruzione contribuisce alla sua estensione e porta ad eventuali fenomeni di aggiustamento qualora il nuovo elemento non sia "coperto" dallo schema.

Il fenomeno della produttività è di tipo scalare ed è fortemente legato agli utilizzi precedenti delle costruzioni. Inoltre, la componente semantica gioca un ruolo fondamentale nel guidare i parlanti ad estendere pattern sintattici a nuovi utilizzi ma non è l'unica componente da prendere in considerazione; sono numerosi i fattori che contribuiscono a rendere produttiva una costruzione e questi riguardano l'aspetto distribuzionale e la frequenza con cui una determinata espressione e i suoi elementi costitutivi ricorrono nella lingua.

Attraverso un riepilogo degli studi più importanti che sono stati compiuti in questo ambito, abbiamo analizzato come i vari fattori interagiscano fra loro e la relativa importanza che ciascuno studioso attribuisce ad ognuno di essi.

3. Obiettivi, materiali e criteri di analisi.

Dopo aver fornito, nei primi due capitoli, una panoramica degli orizzonti teorici entro cui si inserisce l'indagine, questo capitolo è dedicato agli obiettivi prefissi, agli strumenti utilizzati e ai criteri di analisi adottati nel presente studio.

Nel primo paragrafo parlerò degli obiettivi della ricerca, degli studi su cui mi sono basata e che ho già descritto in modo più analitico nei capitoli precedenti, e delle peculiarità della presente indagine rispetto alle altre.

Il secondo paragrafo è dedicato alla scelta dei verbi presi in esame, ai criteri sulla base dei quali sono stati selezionati e al problema circa la definizione del concetto di sinonimia.

Nel terzo paragrafo descriverò gli strumenti utilizzati per estrarre i dati e il corpus preso in esame.

Il quarto e il quinto paragrafo descrivono come sono state compiute l'estrazione e l'analisi dei dati sulla base delle informazioni di frequenza e delle tendenze nella selezione argomentale da parte di ciascuno verbo.

3.1 Obiettivi della ricerca.

Questa indagine si inserisce all'interno di un panorama di studi che hanno come obiettivo quello di indagare la produttività sintattica attraverso un'indagine di carattere sperimentale.

La maggior parte degli studi precedenti (Goldberg, 1995, 2006; Israel, 1996; Bybee e Eddington, 2006; Barðdal, 2008; Matthews e Bannard, 2009; Perek 2015, 2016; e altri) si è soffermata prevalentemente sull'indagine della produttività dello slot verbale all'interno di vari tipi di costruzioni, da quelle di carattere più idiomatico (“the *way* construction”, “the *hell* construction”) a costruzioni più schematiche, analizzando come questo possa essere esteso a nuove forme verbali e cercando di individuare quali sono i fattori responsabili di tali estensioni.

Altri studi si sono invece concentrati sulla produttività dello slot argomentale che segue un verbo (Zeldes, 2013) o un aggettivo (Zeschel, 2010, 2012). È in questa seconda direzione che si è mossa la nostra indagine.

La differenza tuttavia è marginale, entrambi gli approcci partono dall'idea che le costruzioni siano caratterizzate da posizioni libere e variabili (slot) che sono riempite da materiale lessicale nel rispetto dei vincoli imposti dalla costruzione e che la produttività sia un fenomeno scalare e quantificabile.

La nostra ricerca si aggancia ad uno studio compiuto da Zeldes (2012), il quale ha analizzato, a livello statistico, il fenomeno della produttività nel processo di selezione argomentale da parte di gruppi di verbi (quasi) sinonimici o di costruzioni (quasi) sinonimiche.

Alla base dello studio di Zeldes vi è l'intuizione che nel fenomeno della selezione argomentale, agiscano fattori idiosincratici che influenzano la propensione di determinati verbi a essere generalizzati a nuovi argomenti in misura maggiore rispetto ad altri e come tale tendenza non possa essere spiegata facendo ricorso alla sola semantica lessicale.

La spiegazione fornita dallo studioso, in linea con le teorie usage-based del linguaggio, è che la produttività di una costruzione corrisponda al rapporto che intercorre fra il radicamento della costruzione stessa nella mente dl parlante e quello dei suoi argomenti. Nella fattispecie, Zeldes postula che gli elementi più rari contribuiscano maggiormente alla produttività della costruzione in cui ricorrono poiché, essendo essi stessi poco radicati, rafforzano la rappresentazione della costruzione a livello astratto.

The vocabulary size is directly modeled by the members of the network and the probability of novel items corresponds to the proportional entrenchment strength of the construction compared to that of its argument types. Hapax legomena within data for the entire category serve as good estimators of this proportion since they are hardly entrenched themselves and form the greatest contribution to constructional entrenchment that is not biased towards lexicalized prototypes ($V1$ is usually the largest Vm for any m). (Zeldes, 2013: 16)

Lo scopo del presente studio è quello di verificare se le stesse conclusioni sono valide anche per la lingua italiana; ovvero se è possibile delineare una tendenza idiosincratica nella propensione di coppie di verbi quasi sinonimici ad accettare nuovo materiale lessicale nello slot argomentale, che vada al di là di una spiegazione di carattere puramente semantico-lessicale; ci siamo domandati cioè, se all'interno di determinate coppie quasi-sinonimiche uno dei due verbi fosse più predisposto a essere generalizzato a nuovi argomenti rispetto all'altro sulla base della stima del vocabolario con cui ricorre.

La presente analisi condivide con il lavoro di Zeldes l'idea che nel fenomeno della produttività oltre che a fattori di tipo semantico, intervengano fattori idiosincratici propri dell'organizzazione della lingua e che un ruolo di primaria importanza sia svolto dall'esperienza che i parlanti hanno con quest'ultima.

Entrambe le analisi si caratterizzano per la natura corpus-based, i dati, in entrambi i casi, sono stati estratti da corpora linguistici. Il vantaggio dell'utilizzo dei corpora risiede nel fatto che questi ultimi consentono di costruire un modello attendibile, perché fondato su dati "reali", di come si comportano certi fenomeni linguistici, e allo stesso tempo forniscono stime numeriche circa la loro frequenza di ricorrenza, permettendoci di quantificarne la portata e l'importanza.

Per quanto riguarda l'analisi dei dati, Zeldes, riprendendo gli studi che in campo morfologico sono stati compiuti da Bayeen, utilizza essenzialmente misure statistiche come il rapporto (P) fra il numero di *hapax legomena* ($V1$) e il numero totale di ricorrenze all'interno del corpus ($N(C)$), la misura della crescita del vocabolario (VGC) e il limite asintotico S, ovvero il valore massimo di V che ci aspetteremmo di trovare se continuassimo a leggere un testo dello stesso genere per un tempo illimitato; tali misurazioni costituiscono quello che Zeldes definisce *Productivity Complex*.

Nello studio presente, i dati sono stati sottoposti al calcolo della frequenza token per ciascun pattern V+N, della frequenza type, conteggiando il numero dei diversi argomenti con cui ricorre ciascun verbo, e del numero di hapax legomena, ovvero degli argomenti che ricorrono con il verbo con frequenza pari a 1, infine, per ognuno di essi è stata stimata la TTR (type/token ratio) come indice di ricchezza e varietà lessicale.

Le coppie di verbi sono state poi confrontate per verificare se ci fossero delle differenze nella selezione degli argomenti con cui ciascuno di essi ricorre.

Sulla base di questi dati, sono stati creati gli stimoli per un test psicolinguistico realizzato utilizzando una metodologia crowdsourcing. Ai parlanti, madrelingua italiani, sono state sottoposte alcune espressioni composte da V(all'infinito)+N e caratterizzate da pattern di frequenza differenti, ed è stato chiesto loro di esprimere un giudizio di accettabilità.

3.2 La scelta dei verbi.

Per misurare il fenomeno della generalizzazione nella selezione argomentale, abbiamo scelto di utilizzare coppie di verbi quasi-sinonimi, ovvero verbi che, a livello semantico, non presentassero notevoli differenze di significato.

Prima di parlare dei verbi scelti per l'analisi e dei criteri seguiti nella loro selezione, vorrei aprire una breve parentesi circa i dibattiti che sono sorti riguardo al concetto di sinonimia in linguistica e alla sua difficile definizione in termini assoluti.

3.2.1 Il problema della sinonimia in linguistica.

Il concetto di sinonimia costituisce, ancora oggi, una delle principali fonti di dibattito in ambito linguistico. Nonostante, infatti, rappresenti una relazione semantica familiare alla maggior parte dei parlanti, la sua definizione formale non è così semplice.

Comunemente, due parole sono dette “sinonimiche” se hanno “lo stesso significato”, spostando la discussione su un altro grande dibattito circa la nozione di “significato” e la sua altrettanto complessa definizione.

Secondo Leibniz i sinonimi sono parole che possono essere sostituite l'una con l'altra all'interno di una frase senza alterarne il significato (ovvero le condizioni logiche per cui una frase è detta vera o falsa).

Esiste tuttavia una lunga tradizione di studi linguistici, le cui basi risalgono a Bloomfield (1935), in ambito strutturalista, per cui non esisterebbero sinonimi nelle lingue naturali (“there are no synonyms whatsoever in natural language”, Bloomfield, 1935:145), nel senso appena delineato di unità interscambiabili in ogni contesto *salva veritate*.

Affinché una tale equivalenza di significato si verifichi, dovrebbe attuarsi contemporaneamente su più piani: a livello referenziale, a livello distribuzionale, a livello segnico (avere gli stessi tratti semantici, sul piano denotativo e connotativo), e grammaticale (svolgere a tale livello la stessa funzione). A tali condizioni si potrebbero aggiungere anche una similarità nella frequenza di occorrenza e una similarità nei registri di utilizzo (Murphy, 2003).

Una tale coincidenza di fattori è molto difficile che si realizzi.

The first principle of semantic analysis of lexical items is that there are “no synonyms” in the sense that no two lexical items ever have completely the same meanings in all of the contexts in which they might occur. (Louw e Nida, 1988, in Zeldes, 2012:155)

L'unico ambito in cui è possibile trovare una sinonimia stretta è all'interno di linguaggi tecnico-scientifici o con parole dall'estensione di significato molto ristretta (si veda casi come *leone* e *panthera leo* per indicare il medesimo felino).

Al di là di questi casi eccezionali, la perfetta sinonimia diventa una pura questione teorica², ed è per questo motivo che è stata spesso sostituita con la nozione di quasi-sinonimia.

Near-synonyms are words that are almost synonyms, but not quite. They are not fully intersubstitutable, but vary their shades of denotation or connotation, or in the components of meaning they emphasize; they may also vary in grammatical or collocational constraints. (Inkpen e Hirst, 2006: 223)

I quasi-sinonimi possono variare in molteplici modi: a livello denotativo, a livello di frequenza di utilizzo e di *attitude* che il parlante ha nei confronti dell'entità denotata (attribuendo un valore peggiorativo, migliorativo o neutrale), a livello stilistico per quanto riguarda la formalità, la concretezza e la familiarità con cui un termine è utilizzato, e infine, possono presentare vincoli collocazionali differenti.

Questo fatto tuttavia non preclude la possibilità che determinate classi semantiche vadano a disegnare uno spazio semantico all'interno del quale verbi quasi-sinonimi selezionano i propri argomenti.

Sulla base di questi assunti, per condurre la nostra analisi sono state scelte venti coppie/terne di verbi quasi-sinonimi, prendendo tale definizione con tutte le cautele del caso.

² Un contributo importante, per quanto riguarda il concetto di sinonimia e, più in generale, quello di similarità, è costituito dalla *semantica distribuzionale*, la quale, sulla base delle parole del linguista inglese J.R. Firth, “you shall know a word by the company it keeps” (1957:11), assume che la distribuzione delle parole nei contesti linguistici giochi un ruolo chiave nel caratterizzarne il comportamento semantico. Tale ipotesi trova fondamento in quella che viene definita come *Weak Contextual Hypothesis*, per cui, la somiglianza della rappresentazione contestuale di due parole contribuisce a determinarne il grado di similarità.

3.2.2 Criteri di selezione dei verbi.

Per condurre la nostra indagine abbiamo selezionato venti coppie/terne di verbi quasi-sinonimi, presi da vari campi semantici, alcuni caratterizzati da un significato più ristretto, altri da uno più generale. La scelta dei verbi è avvenuta per lo più senza consultare preventivamente i dati forniti dal corpus ma in base a ciò che, a livello intuitivo, si riteneva potesse costituire un caso interessante di indagine.

Sono stati scelti verbi che non ricorressero all'interno del corpus con una frequenza eccessivamente elevata, per una questione di maggiore maneggevolezza dei dati durante la fase di analisi, né con una frequenza troppo bassa.

Delle venti coppie/terne di verbi, 18 sono costituite da verbi transitivi, e 2 da verbi intransitivi. Tale distinzione non è dovuta a particolari esigenze d'indagine.

Figura 9 Tabella dei verbi scelti per l'indagine senza consultare il corpus.

Verbi quasi-sinonimi
Inghiottire-ingoiare
Incappare in- incorrere in
Rompere-infrangere
Estirpare-sradicare
Vietare-proibire
Spaventare-impaurire
Imparare-apprendere
Detestare-odiare-aborrire
Criticare-deplorare-deprecare
Gridare-urlare
Finire-terminare
Colmare-riempire
Rinnegare-disconoscere-ripudiare
Calmare-placare-quietare
Comprare-acquistare
Adempiere a-ottemperare a-assolvere a
Temere-paventare
Scordare-dimenticare
Guidare-condurre
Cominciare-iniziare-avviare

In generale, i verbi scelti appartengono ad ambiti semantici differenti e si distinguono anche per la frequenza con cui ricorrono all'interno della lingua.

Alcuni di essi vanno a delineare campi semantici più ristretti e specifici (*adempiere, ottemperare, assolvere, rinnegare, ripudiare, sradicare, estirpare*) e questo si riflette anche nella gamma di argomenti con cui i verbi ricorrono, che sarà più ristretta e circoscritta a livello semantico, altri invece delineano azioni o stati più generali e ricorrono con un numero maggiore e vario di elementi lessicali.

Inoltre, alcuni sono di uso più comune (*cominciare, finire, dimenticare, comprare*), altri invece sono limitati a determinati registri linguistici (*aborrire, paventare, quietare*).

Pattern di frequenza differenti sono stati postulati anche all'interno di alcune coppie (temere-paventare, odiare-aborrire, criticare-deplorare), in linea con quanto abbiamo detto circa la nozione di quasi-sinonimia, per l'appartenenza dei verbi a registri differenti (temere-paventare), o perché riflettono *attitude* diversi che i parlanti hanno nei confronti dell'azione denotata, per esempio *aborrire* e *detestare*, il primo più formale del secondo, esprimono entrambi un sentimento più intenso dell'odio.

3.2.3 Strumenti per la selezione dei verbi.

Per la scelta dei verbi mi sono avvalsa del vocabolario dei sinonimi e contrari Garzanti (2009), del vocabolario on-line dei sinonimi e contrari della Treccani (<http://www.treccani.it>), della versione on-line del dizionario dei sinonimi e contrari disponibile sul sito del Corriere della Sera (<http://dizionari.corriere.it>), e del sito della Accademia della Crusca (<http://www.accademiadellacrusca.it>). In alcuni casi è stata utilizzata la funzione *sketch-difference* disponibile su Sketch Engine (<https://www.sketchengine.co.uk>), uno strumento di analisi dei corpora che prende come input un corpus e dei pattern grammaticali corrispondenti. Tale funzione mostra somiglianze e differenze tra quasi-sinonimi specificando, per due parole legate semanticamente, quale comportamento condividono e in cosa differiscono. Questo strumento si è rivelato particolarmente utile per verificare il grado di sovrapposizione semantica di due parole ma anche le differenze di selezione.

3.3 Strumenti di indagine e il corpus di riferimento.

Ogni verbo è stato estratto insieme agli argomenti che ricorrevano, nel corpus, nella posizione di oggetto diretto per i verbi transitivi e oggetto indiretto per quelli intransitivi.

Come strumento d'indagine e come mezzo per l'estrazione dei dati è stato utilizzato Lexit, un framework computazionale per l'acquisizione automatica e l'esplorazione di informazioni distribuzionali riguardo a nomi, verbi e aggettivi italiani che fa capo a due corpora: il corpus La Repubblica e la sezione italiana di Wikipedia.

3.3.1 Lexit.

I dati sono stati estratti con l'ausilio di Lexit (<http://lexit.fileli.unipi.it>), un sito che permette l'esplorazione dei profili distribuzionali di nomi, aggettivi e verbi italiani. Tali profili, contengono un ampio repertorio d'informazioni statistiche estratte automaticamente da corpora con metodi linguistico-computazionali. Attraverso gli indicatori statistici associati a ciascun tipo d'informazione, è dunque possibile analizzare le proprietà combinatorie delle parole e la struttura argomentale dei predicati identificandone i tratti distribuzionali più salienti e prototipici.

All'interno di Lexit, ogni lemma target è associato a un profilo distribuzionale, ovvero un insieme di informazioni statistiche estratte da un corpus, in questo caso La Repubblica, applicando le ultime metodologie di NLP, senza revisione manuale.

Tali informazioni sono articolate in due profili, un profilo sintattico, che specifica lo slot sintattico e il frame di sotto-categorizzazione con cui ricorre il predicato, e un profilo semantico, costituito dall'insieme lessicale dei *fillers* più prototipici che realizzano lo slot sintattico e dalle classi semantiche che ne caratterizzano le preferenze di selezione.

Per identificare i tratti più salienti delle strutture argomentali, ogni caratteristica distribuzionale è pesata con la frequenza di ricorrenza nel corpus e con una stima della *Local Mutual Information* (LMI), una variante della *Pointwise Mutual Information*, utilizzata in Lexit per misurare l'associazione fra il verbo e il frame di sotto-categorizzazione, fra lo slot e i *fillers* lessicali e, infine, fra lo slot e le classi semantiche.

La versione attuale di Lexit contiene informazioni raccolte da due corpora differenti: il corpus *La Repubblica*, una collezione di articoli di giornale costituita da circa 380 milioni di tokens, e la sezione italiana di Wikipedia (152 milioni di tokens.). I corpora sono tokenizzati, lemmatizzati e analizzati morfosintatticamente con TANL (*Text Analytics and Natural Language*), un insieme di moduli di NLP per la lingua italiana, sviluppato presso l'Università di Pisa in collaborazione con ILC-CNR.

L'interfaccia web consente agli utenti di selezionare il lemma target desiderato (verbo, nome, aggettivo) e il corpus in cui compiere l'indagine.

La possibilità di combinare diversi parametri di ricerca rende questo strumento molto funzionale e flessibile per affrontare diversi tipi di ricerche in ambito linguistico computazionale e lessicografico, inoltre l'interfaccia grafica facilita la ricerca rendendola di facile accesso all'utenza.

3.3.2 Il corpus La Repubblica.

Il corpus che abbiamo selezionato come fonte per i nostri dati è La Repubblica, un corpus realizzato al SSLMIT dell'Università di Bologna. Costituisce una vasta collezione di testi giornalistici che ammontano approssimativamente a 380 milioni di tokens e contiene articoli pubblicati fra il 1985 e il 2000, provenienti da uno dei quotidiani più letti dalla popolazione italiana: La Repubblica.

Il corpus è *tokenizzato*, analizzato morfosintatticamente, lemmatizzato e categorizzato per genere e topic.

Tale risorsa costituisce uno strumento fondamentale per condurre analisi linguistiche sulla lingua italiana contemporanea poiché è probabilmente l'unico corpus di tali dimensioni a essere accessibile a tutti.

È tuttavia importante sottolineare la presenza di alcuni punti deboli. Per prima cosa, tutti i testi sono istanza dello stesso macro-genere, quello giornalistico, e provengono dalla medesima fonte. Pertanto, per quanto il compito del corpus sia quello di approssimare l'esperienza linguistica dei parlanti, è bene tenere conto che tale modello sarà rappresentativo di quel tipo di registro o testo. Inoltre, il corpus è stato processato e annotato automaticamente, pertanto, molti errori di battitura, segmentazione, parsing e certe idiosincrasie non sono state eliminate. Infine, le informazioni extra-testuali e strutturali sono scarse; non è possibile né cercare due parole che appaiono nello stesso

paragrafo né i testi che compaiono nella cosiddetta “terza pagina”, ovvero quella sezione che solitamente è dedicata alla cultura e alle arti.

3.4 Estrazione dei dati da Lexit.

La ricerca su Lexit è stata condotta selezionando come lemma il verbo sul quale si voleva compiere l’analisi distribuzionale, e ricercando i *fillers* dello slot dell’OBJ, nel caso dei verbi transitivi, e dei complementi indiretti d’interesse per i verbi intransitivi.

Oltre ai pattern V+N sono state estratte, dal corpus La Repubblica, le frasi in cui essi ricorrevano all’interno del corpus, selezionando una finestra contesto di 20 parole a destra e 20 parole a sinistra del lemma ricercato. Tale ampiezza è dovuta alla volontà di analizzare correttamente i dati in modo da eliminare i casi non rappresentativi del pattern preso in analisi.

I dati sono stati estratti insieme alle informazioni di frequenza.

Prima dell’analisi delle coppie, tuttavia è stata necessaria un’operazione di ripulitura manuale dei dati. Ho ritenuto necessaria questa fase, dal momento che il corpus è processato e annotato automaticamente, per essere sicura che il nome che ricorreva nello slot argomentale fosse occupato effettivamente dall’oggetto diretto o indiretto del verbo, e per evitare errori di varia natura come ripetizioni dovute a varianti grafiche (es. week-end, weekend; tv, televisione, tivù), errori di battitura, errori di *tagging*, ecc.. Per compiere tale operazione è stato necessario analizzare ogni pattern V+N all’interno del contesto-frase in cui ciascuno di essi ricorreva nel corpus.

L’analisi manuale è stata condotta sugli oggetti di ciascun verbo, con particolare attenzione ai pattern V+N che nel corpus avevano un numero di occorrenze totali, basso. In questi casi, il peso dell’errore è maggiore rispetto che nei pattern ad alta frequenza, perché a causa della scarsità dei dati, gli errori non hanno la possibilità di essere riassorbiti e attenuati.

In alcuni casi, tale analisi si è rivelata di vitale importanza.

Con i verbi dichiarativi (*gridare-urlare*) sono stati costatati numerosi errori di *tagging*: spesso, quello che in realtà è il soggetto del verbo, era stato riconosciuto come oggetto (casi come *bambino, leader, folla, donna, uomo*), soprattutto se posposto rispetto al verbo all’interno della frase.

21) Andate via, noi vi bruceremo, urlava la folla. [La Repubblica, 3622255]

22) Il principe, grida un bambino, voglio il principe e batte le mani. [La Repubblica, 1760709]

In questi casi, la stima del vocabolario del verbo e del numero di hapax legomena, è variata radicalmente dopo l'analisi, come si può vedere nella tabella sotto riportata.

Figura 10 Analisi manuale degli errori rispetto ai dati estratti automaticamente per la coppia *gridare- urlare*.

verbo	V_(automatico)	V1_(automatico)	errori	V_(manuale)	V1_(manuale)
gridare	1262	782	899	363	201
urlare	866	592	601	265	171

La tipologia degli errori incontrati ha riguardato, per la maggior parte, errori di *tagging*: soggetti, complementi tempo (*ora, anno, mese*), quantificatori o espressioni analoghe (*metro, chilometro, ettaro* ma che espressioni come *specie di, sorta di*) riconosciuti come oggetti.

23) Jil vuol fare il colpo: comprò lo scorso anno Futre dal Porto per cinque miliardi, ora vuole rivenderlo a dieci. [La Repubblica, 2511579]

Un caso particolare è rappresentato dai verbi *cominciare-iniziare* e *terminare-finire-ultimare*. Questi ultimi, infatti, hanno un comportamento sia transitivo (*i ragazzi cominciarono le vacanze estive*) che intransitivo (*le vacanze iniziarono ad agosto*). Quando intransitivi, essi si presentano come *inaccusativi*, una classe di verbi che si caratterizza, in prima istanza, per la presenza dell'ausiliare *essere* nei tempi composti (*le vacanze sono iniziate ad agosto*)³.

Secondo la formulazione originaria dell'*ipotesi inaccusativa*, il soggetto dei questi verbi sarebbe *soggetto* a livello superficiale, ma *oggetto* a livello sottostante.

Per quanto riguarda il presente lavoro, numerosi erano i casi nel corpus in cui quello che era il soggetto di un verbo inaccusativo era stato riconosciuto come oggetto. In questa

³ Oltre a questo, presentano una serie di peculiarità che ne permettono la riconoscibilità, fra cui la pronominalizzazione del soggetto con il clitico partitivo *ne*, il fatto che, al participio passato, essi possono apparire in costruzione assoluta con il soggetto posposto (*terminato il lavoro, Luca tornò a casa*), e che il participio passato può agire come modificatore aggettivale (per maggiori approfondimenti cfr. Bentley, 2006; Levin e Rappaport Hovav, 1995; Zaenen, 2006).

particolare circostanza, il modus operandi adottato, è stato quello di accettare anche questi casi nel rispetto dell'ipotesi inaccusativa. Sono numerosi, inoltre, i casi in cui lo stesso nome talvolta ricorre come oggetto del verbo transitivo e come soggetto dell'inaccusativo, per cui sarebbe stata impossibile una selezione dei soli casi di tagging corretto e avrebbe causato un grande dispendio di tempo.

Si vedano i casi:

24) Il concerto è terminato dopo la mezzanotte [La Repubblica, 2608468]

25) Daniele, che da un paio di anni ha molto rallentato la sua attività dal vivo, si è poi lanciato in una movimentata “Long train running” dei Doobie Brothers per terminare il suo piccolo concerto con “Imagine” di Lennon. [La Repubblica, 6148402]

Anche casi come (26) e (27) sono stati mantenuti:

26) Il suo amore per i libri è cominciato fin da adolescente. [La Repubblica, 3668335]

27) L'inchiesta è quasi terminata. [La Repubblica 2579186]

I dati, così filtrati a mano, sono stati riordinati in tabelle Excel e per ogni pattern V+obj/V+ind. Obj, è stata calcolata la frequenza token, la frequenza type, il numero di hapax e la TTR(type/token ratio).

3.5 Analisi dei dati estratti.

Nella tabella sotto riportata è possibile analizzare i dati ricavati dall'estrazione per ciascuna coppia di verbi quasi sinonimi. Per condurre le varie misurazioni di frequenza è stato utilizzato Excel, un programma prodotto da Microsoft, dedicato alla produzione e alla gestione di fogli di calcolo. Sulla base dei dati ricavati, in seguito analizzeremo solo i casi che riterremo più interessanti ai fini della nostra indagine e ci concentreremo soprattutto sulle coppie di verbi che abbiamo deciso di utilizzare per la realizzazione dell'esperimento.

Qui sotto riportiamo la lista dei verbi che prenderemo in considerazione nell'analisi⁴.

Figura 11 Dati ricavati da Lexit. Per ciascun verbo sono state riportate le misure di frequenza token, frequenza type, numero di hapax legomena, indice di TTR (type/token ratio).

Verbo	Frequenza token	Frequenza type	Numero hapax	TTR	TTR%
Incorrere in	1141	308	192	0,26993865	26,99%
Incappare in	1185	495	328	0,417721519	41,77%
Inghiottire	732	426	326	0,581967213	58,20%
Ingoiare	1009	487	362	0,482656095	48,27%
Assolvere a	695	71	40	0,102158273	10,22%
Adempiere a	721	90	42	0,12482663	12,48%
Ottemperare a	441	108	63	0,244897959	24,49%
Estirpare	299	145	102	0,484949833	48,49%
Sradicare	489	235	164	0,480572597	48,06%
Scordare	382	295	254	0,772251309	77,23%
Dimenticare	10937	2998	1618	0,274115388	27,41%
Detestare	991	634	465	0,63975782	63,98%
Odiare	1497	770	538	0,496325985	49,63%
Placare	1614	404	243	0,250309789	25,03%
Calmare	1331	307	181	0,230653644	23,07%
Apprendere	1898	374	240	0,197049526	19,70%
Imparare	3130	597	357	0,190734824	19,27%
Colmare	2440	262	172	0,107377049	10,73%
Riempire	6118	1006	545	0,164432821	16,44%
Sollevare	9239	820	443	0,088754194	8,87%
Alzare	12733	723	357	0,056781591	5,67%
Romper	10487	912	441	0,086964814	8,69%
Infrangere	1881	370	226	0,196703881	19,67%
Guidare	14632	1367	605	0,093425369	9,34%
Condurre	11816	902	475	0,07633717	7,63%
Cominciare	14299	1569	685	0,109727953	10,97%
Iniziare	14416	1503	684	0,104259156	10,43%
Avviare	19607	1440	656	0,073443158	7,34%
Comprare	13298	1925	938	0,14483381	14,48%
Acquistare	14359	1651	801	0,114980152	11,50%
Finire	10502	1813	953	0,172633784	17,26%
Terminare	3482	763	418	0,219126939	21,91%

⁴ Per una lista dei verbi estratti completa cfr. appendice.

3.5.1 Frequenza Token, frequenza type, type-token ratio.

In questo paragrafo ci dedicheremo ad analizzare i dati alla luce delle informazioni ricavate dal corpus circa la *frequenza token*, la *frequenza type*, il numero di *hapax legomena* che ricorrono con i verbi e l'indice di *TTR*.

Il primo dato riportato nella tabella indica la *frequenza token*, ovvero il numero delle occorrenze totali dei vari pattern N+V nel corpus. Tale misurazione corrisponde all'indice N(C) di cui parla Zeldes e fornisce un'indicazione circa la frequenza con cui un determinato pattern è utilizzato all'interno di una lingua e quindi del suo grado di radicamento.

Con il termine *frequenza token* ci si riferisce anche alle singole occorrenze di un tipo nel corpus. Tale misura costituisce un indice altrettanto importante poiché riflette la forza del legame fra un verbo e il nome che lo accompagna, e indica, dunque, quanto è radicata una singola espressione e quale sia la sua salienza cognitiva. Come abbiamo visto negli studi precedenti, espressioni caratterizzate da un'elevata *frequenza token*, possono costituire un modello per estensioni di carattere analogico.

La *frequenza type* indica il numero di elementi lessicali distinti che sono istanza di uno slot argomentale. Tale misurazione corrisponde a ciò che Zeldes e Bayeen definiscono con V (*Vocabulary*) e costituisce una misura importante poiché, secondo quanto emerge dagli studi precedenti, i pattern caratterizzati da una *frequenza type* elevata sono solitamente più predisposti ad accettare nuovi elementi lessicali all'interno delle loro posizioni aperte.

Il numero degli *hapax legomena* (V1) misura, invece, le istanze che ricorrono con frequenza pari a 1. Come abbiamo visto nel capitolo precedente, secondo Bayeen e Zeldes, il numero di *hapax legomena* costituisce un indice importante, sulla base dell'assunto euristico secondo il quale gli elementi che, all'interno di un corpus o di un testo, ricorrono una volta sola, costituiscono neologismi, ovvero nuove attestazioni di un pattern⁵.

⁵ I fatti, ovviamente, non rispecchiano sempre questa condizione. Numerosi sono i casi in cui determinati elementi ricorrono in un corpus con frequenza pari a 1 per pura casualità, nonostante siano forme lessicalizzate e soprattutto familiari alla maggior parte dei parlanti. Ciò non toglie, tuttavia, che V1 costituisca un'interessante categoria di indagine e che possa fornire una qualche idea circa la probabilità che ha un determinato pattern di essere esteso a nuove forme non attestate.

Infine la *type-token ratio* (TTR) è un indice della creatività e della ricchezza lessicale di un testo poiché ne indica la variazione del vocabolario.

Si potrebbe affermare che la TTR indica, in valori percentuali, il numero di argomenti diversi che ricorrono insieme a un determinato verbo, in rapporto al numero complessivo di argomenti che ricorrono con lo stesso.

Anche in questo caso, è opportuno compiere una precisazione dovuta al fatto che la TTR dipende dalla lunghezza del testo: più un testo è lungo, più questo rapporto è destinato a calare rapidamente, dato che il numero di parole nuove cresce a un ritmo minore di quanto aumenti la frequenza delle parole che si ripetono (si veda a tale riguardo la differenza del rapporto fra verbi a bassa frequenza come *aborrir* e *deprecare*, *scordare*, *impaurire*, che presentano indici molto elevati di TTR e verbi ad alta frequenza come *temere*, *cominciare*, *terminare*, *finire* ecc., che riportano valori più bassi)⁶. Nonostante, dunque, alcune misurazioni siano poco informative a causa della scarsa numerosità del campione, nel complesso tale indice costituisce un indicatore importante della varietà lessicale degli elementi che vanno a occupare lo slot argomentale dei pattern presi in analisi.

Per quanto riguarda i dati estratti, possiamo notare che i pattern tendono a essere collocabili più o meno distintamente su tre fasce di frequenza: alta frequenza (*finire*, *dimenticare*, *acquistare*, *comprare*, *cominciare*, *iniziare*, *avviare*, *temere*, *condurre*, *guidare*, ecc.), media frequenza (*terminare*, *imparare*, *apprendere*, *colmare*, *gridare*, *proibire*, ecc.), bassa frequenza (*ottemperare a*, *assolvere a*, *adempiere a*, *impaurire*, *inghiottire*, *sradicare*, *estirpare*, ecc.).

La frequenza dei pattern, all'interno del corpus, riflette la frequenza con cui questi ultimi sono utilizzati nella lingua di tutti i giorni. Determinate espressioni, infatti, sono utilizzate più frequentemente di altre nella comunicazione quotidiana. Tale tendenza rispecchia la distinzione che aprioristicamente era stata ipotizzata: i verbi che ricorrono nel corpus con frequenza più elevata sono anche quelli che sono utilizzati quotidianamente con frequenza maggiore (*comprare*, *acquistare*, *cominciare*, *iniziare*, *terminare*, *finire*, *vietare*, *dimenticare*, *imparare*, ecc.).

Inoltre possiamo notare come all'interno di determinate coppie di quasi-sinonimi uno dei due verbi ricorra nel corpus, accompagnato da argomento diretto/indiretto, con una frequenza maggiore rispetto all'altro: questo è il caso di *odiare* rispetto a *detestare* e

⁶ Per venire meno a questa problematica, è stata spesso sostituita dalla STTR (*Standardised Type/Token Ratio*) che misura lo stesso valore però calcolato su un campione di *n* tokens.

aborrire, *temere* rispetto a *paventare*, *dimenticare* rispetto a *scordare*, *criticare* rispetto a *deplorare*, *spaventare* rispetto a *impaurire*, rispecchiando spesso differenze di registro nell'utilizzo dell'uno rispetto all'altro.

Interessante il caso della coppia *scordare-dimenticare*. I due verbi costituiscono un caso di sinonimia quasi completa, infatti diciamo con la stessa naturalezza di aver dimenticato o scordato qualcosa o qualcuno secondo le diverse accezioni presentate dai due verbi (“perdere il ricordo di qualcosa o di qualcuno”, “privare qualcuno del proprio affetto”, “lasciare qualcosa in un posto per disattenzione”, “non tenere in considerazione, non fare caso”), inoltre i due verbi non sono caratterizzati da una distinzione di carattere regionale o di registro: occorrenze di entrambi i termini sono registrate su tutto il territorio italiano, anche se con una distribuzione geografica disomogenea, con eccezione del sardo che utilizza la forma *iskarèssere* o *scarèširi* che deriva dal latino volgare *excadere (Meyer-Lubke, *Romanisches Etymologisches Wörterbuch*, 1935: 261).

L'uso tuttavia sancisce una differenza non trascurabile delle due forme che è stata registrata anche da GRADIT⁷ il quale classifica *dimenticare* come lemma fondamentale⁸ del vocabolario di base e *scordare* come di alto uso. Tale differenza si rispecchia nei dati ricavati dal corpus in cui il verbo *dimenticare*, accompagnato dall'oggetto diretto, ricorre con una frequenza token pari a 10937, mentre *scordare*, con frequenza molto più bassa (382 tokens).

Inoltre, possiamo notare come entrambi siano caratterizzati da un'elevata frequenza type e da un elevato numero di *hapax legomena*, che, seguendo Bayeen e Zeldes, sono un ulteriore indicatore della predisposizione di un verbo a essere esteso a nuovi utilizzi. Dei verbi ad alta frequenza, *dimenticare* è quello che presenta il vocabolario più ampio, allo stesso modo, *scordare*, sebbene abbia una frequenza molto più bassa, presenta, rispetto ai verbi che appartengono alla medesima fascia, una frequenza type elevata⁹.

⁷ Il Grande Dizionario Italiano dell'Uso, è un dizionario curato da Tullio De Mauro. Con 270000 vocaboli in sei volumi più due supplementi (2003 e 2008) è il più esauriente lemmario della lingua italiana dell'uso corrente.

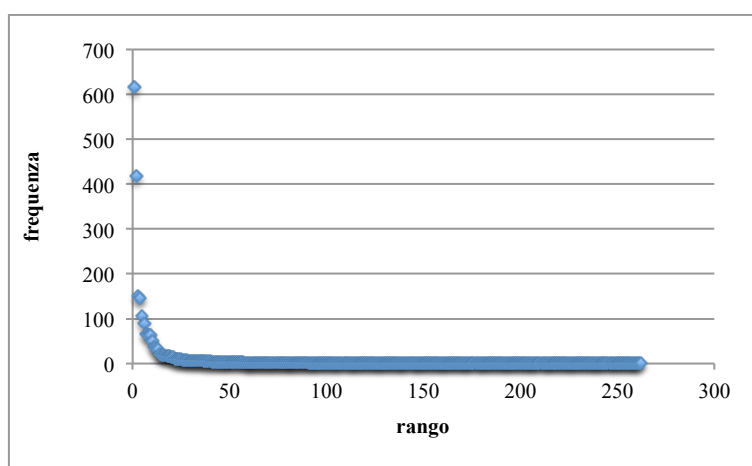
⁸ Si ricorda che con l'etichetta FO, con cui sono marcati i lemmi fondamentali, si fa riferimento a circa 2000 termini che compongono il nostro lessico quotidiano, mentre AU, identifica 2500 termini che sarebbero utilizzati solo nel 6% dei nostri discorsi. Tale distinzione è dunque di rimarchevole importanza per quanto riguarda la nostra indagine.

⁹ Si veda, per esempio, come *scordare*, presenti un numero di types di poco minore rispetto a *incorrere*, nonostante quest'ultimo abbia una frequenza token molto più elevata. Lo stesso si può dire rispetto a *colmare* che attesta 262 types nella posizione di complemento oggetto su un numero di occorrenze totali pari a 2440 tokens.

Tale ricchezza lessicale trova riscontro anche negli indici elevati di TTR (*scordare*: 77,23%; *dimenticare*: 27,41%), come riflesso di una certa varietà e ricchezza lessicale. Una coppia interessante è rappresentata dai verbi *riempire-colmare*. Come possiamo vedere dalla tabella, i pattern rappresentati da *riempire+N* e *colmare+N*, si presentano nel corpus con frequenze nettamente distinte, con una presenza più massiccia del primo. Nette anche le distinzioni che riguardano il vocabolario: *riempire* ricorre con un vocabolario più ampio rispetto a *colmare*, e ciò si riflette negli indici di TTR (*riempire+N*, 16,44%; *colmare+N*, 10,73%). In generale, rispetto ad altri pattern che presentano frequenze di ricorrenza analoghe, si nota come nessuno dei due verbi sia comunque caratterizzato da un vocabolario particolarmente ricco e vario. Inoltre, per quanto riguarda *colmare*, se analizziamo la frequenza di ricorrenza dei nomi che compaiono nei primi ranghi, in un ordine discendente di frequenza, notiamo come il verbo ricorre con frequenze molto elevate con i primi venti nomi, con i quali stabilisce legami forti (es. *colmare+lacuna*: 418 tokens, *colmare+vuoto*: 616 tokens), mentre piano piano che si scende di rango, notiamo che le frequenze delle singole espressioni subiscono un calo molto repentino (cfr. paragrafo successivo per analizzare le preferenze di selezione dei due verbi).

Come si può vedere nel grafico riportato sotto, la distribuzione delle frequenze degli argomenti di *colmare*, rispetto al loro rango, segue un andamento tipicamente zipfiano¹⁰.

Figura 12 Rapporto tra frequenza degli argomenti di *colmare* e il loro rango.



¹⁰ La legge di Zipf (1949) stabilisce che la frequenza di una parola all'interno di un testo è inversamente proporzionale al suo rango: $f(z) = C/z^a$. Ne deriva che all'interno di un testo, ci saranno poche parole molto frequenti, e molte parole a bassa frequenza e hapax: “*V1* is usually the largest *Vm* for any *m*” (Zeldes, 2012:16).

Una situazione analoga si verifica con la coppia *rompere-infrangere*. La frequenza con cui ricorre il pattern *rompere+N* rispetto a *infrangere+N* è molto più alta (10487 tokens vs. 1881 tokens). Entrambi i verbi presentano un vocabolario abbastanza ridotto, e questo si riflette in indici di TTR abbastanza bassi (8,69% *rompere+N*; 19,67% *infrangere+N*) rispetto anche ad altri verbi che si collocano su fasce di frequenza simili o inferiori. Si pensi che *dimenticare*, che ha una frequenza token simile (pari a 10937 tokens), presenta un vocabolario più di tre volte più ampio rispetto a *rompere*; tale andamento riflette, probabilmente, il fatto che il numero di cose, concrete e astratte, che si possono rompere è minore rispetto al numero delle cose che possono essere dimenticate. Tuttavia è interessante notare come *infrangere*, nonostante non abbia frequenza particolarmente elevata, abbia un indice di TTR non bassissimo, pari a 19,67%.

Sollevare-alzare, pur ricorrendo nel corpus con frequenze abbastanza elevate, presentano un vocabolario ridotto che si riflette in bassi valori di type/token ratio: *sollevare+N*, 8,87% e *alzare+N* 5,67%. Una differenza interessante emerge però se analizziamo le differenze che intercorrono fra i due verbi nella coppia.

Alzare+N, ricorre nel corpus con frequenza maggiore rispetto a *sollevare+N*, però il secondo presenta un vocabolario più ampio e anche un numero di hapax legomena più elevato. Pertanto, l'indice di TTR risulta più elevato per il secondo pattern (*sollevare+N*: 8,87% vs. *alzare+N*: 5,67%), lasciando presagire che il verbo sia caratterizzato da una maggiore varietà e ricchezza lessicale.

Un caso opposto è costituito dalla coppia *odiare+N*, 1497 tokens, e *detestare+N*, 991 tokens.

Entrambi sono caratterizzati da un vocabolario particolarmente numeroso e da un elevato numero di hapax legomena. Questo è dovuto al fatto che la classe delle cose che posso essere odiate o detestate sono tante e appartengono a classi semantiche distinte (*person, animal, substance, food, knowledge, feeling, artifact, natural-object, plant*)¹¹. Tale ricchezza lessicale si traduce in una stima elevata di TTR per entrambi i pattern (63,98% *detestare+N*; 49,63% *odiare+N*). Sebbene i due verbi ricorrano con frequenze differenti, è tuttavia interessante notare come *detestare* ricorra comunque in proporzione con un numero maggiore di argomenti rispetto a *odiare*, lasciando intuire la presenza di un vocabolario più ricco.

¹¹ Classi semantiche degli argomenti che ricorrono con *odiare* e *detestare*, secondo Lexit (<http://lexit.fileli.unipi.it/>).

Un altro caso interessante è rappresentato dalla coppia *comprare-acquistare*. Il pattern *comprare+N* ricorre con una frequenza meno elevata rispetto ad *acquistare+N* (13298 tokens vs. 14359 tokens)¹², tuttavia il primo presenta un indice di ricchezza lessicale più elevato perché proporzionalmente ricorre con un maggior numero di nomi differenti: la TTR di *comprare+N* equivale a 14,48% mentre quella di *acquistare+N* è pari a 11,50%. La stessa differenza si può notare con il numero di hapax legomena: *comprare*, presenta un numero di argomenti con frequenza pari a 1, superiore rispetto ad *acquistare*. A partire da questi dati, e sulla base di quanto asserito da Zeldes e da Bayeen negli studi precedenti, dunque si potrebbe ipotizzare che *comprare* abbia una maggiore predisposizione a essere generalizzato a nuovi usi e nuovi argomenti.

Un caso analogo, su fasce di frequenza molto più basse, è rappresentato da *inghiottire-ingoiare*. Anche in questo caso il pattern *inghiottire+N* ricorre nel corpus con una frequenza più bassa rispetto a *ingoiare+N* di circa 300 tokens, mentre i due verbi presentano una differenza di vocabolario abbastanza marginale (426 types vs. 487 types). Pertanto, il primo è caratterizzato da un valore di TTR più elevato (58,20%) rispetto al secondo (48,27%) che porta a ipotizzare una sua maggiore predisposizione ad accogliere nuovi elementi lessicali all'interno dello slot argomentale.

Un caso opposto è rappresentato dai verbi *incorrere-incappare*. La coppia non presenta differenze particolarmente significative di frequenza token (*incappare in* N(C)= 1185; *incorrere in* N(C)=1141), mentre si distinguono soprattutto per quanto riguarda il vocabolario (V), *incorrere in* ricorre con 308 types, *incappare in* con 495, e per il numero di hapax, maggiore nel caso del secondo verbo (328 hapax, *incappare* vs. 192 hapax, *incorrere*). Da questo ne deriva una differenza anche di TTR, pari a 26,99% nel primo caso e 41,77% nel secondo, da cui si deduce che *incappare* presenta un vocabolario più vario e ricco rispetto a *incorrere*.

Anche in questo caso, i due verbi presentano una diversa distribuzione degli argomenti. Nel caso di *incorrere*, gli argomenti appartenenti ai ranghi più alti, sempre in un ordine di frequenza decrescente, hanno un maggiore radicamento rispetto agli argomenti più frequenti di *incappare* che presentano, in generale frequenze più basse e una

¹² Notare che *acquistare*, presenta dei significati secondari che non condivide con *comprare*: “procurarsi, procacciarsi” nel senso di *acquistare fama*, *acquistare esperienza*, *acquistare forza*, “guadagnare” come *acquistare terreno*, *acquistare tempo*, *acquistare altezza*, e “aggiungere”, *acquistare merito*, *acquistare fama*, *acquistare valore*. I nomi selezionati dal verbo quando assumeva questi significati, non sono stati conteggiati nel calcolo delle frequenze e della TTR.

distribuzione più omogenea: l'espressione *incorrere in+sanzione* (rango 1) ricorre 134 volte vs. *incappare+maglia* (rango 1), 60 tokens.

Nella coppia *imparare-apprendere*, si verifica una situazione differente. I due pattern ricorrono nel corpus con frequenze differenti (*imparare+N* ricorre 3130 volte mentre *apprendere+N*, 1898), e presentano un vocabolario di dimensione diversa (597 types *imparare*, e 374 types *apprendere*) tuttavia la stima della TTR è molto simile (19,70% *apprendere*, 19,27% *imparare*), lasciando presagire una ricchezza e una varietà lessicale simili.

Con *estirpare-sradicare*, si presenta una situazione analoga anche se su fasce di frequenza molto più basse. Il pattern *sradicare+N* ricorre con frequenza più elevata rispetto a *estirpare+N* (299 tokens vs. 489 tokens). Se però li andiamo ad analizzare notiamo che essi hanno un comportamento analogo, ovvero, il numero di types con cui ricorre ciascun verbo è circa la metà delle occorrenze totali dell'espressioni. Ne risulta, una stima analoga di TTR (*estirpare+N*, 48,49% e *sradicare+N*, 48,06%) per cui è ipotizzabile che i due verbi abbiano un indice simile di ricchezza lessicale.

I pattern costituiti da *calmare+N* e *placare+N*, non presentano differenze molto nette nella frequenza di ricorrenza nel corpus. *Placare* ricorre con frequenza leggermente più elevata, e in proporzione presenta un vocabolario più ampio e un numero di hapax legomena più alto; il suo indice di TTR è leggermente più elevato rispetto a quello di *calmare*.

Condurre+N e *guidare+N*, sono rappresentati rispettivamente da 11816 tokens il primo e 14632 tokens il secondo. Non presentano un vocabolario particolarmente ricco rispetto a verbi come *comprare* e *dimenticare*, aspetto che si riflette in valori non alti di TTR (7,63% il primo, e 9,34% il secondo). Insieme ad *alzare+N* e *sollevare+N*, costituiscono i pattern caratterizzati dalla minore varietà lessicale nel vocabolario. Anche in questo caso la differenza può essere spiegata con il fatto che la classe delle cose che possono essere condotte o guidate, è sicuramente meno numerosa rispetto a quella delle cose che possono essere comprate o dimenticate.

Un gruppo di verbi interessanti è rappresentato da *avviare-cominciare-iniziare*. Sempre facendo riferimento alla [tabella 3](#) possiamo vedere come *iniziare* e *cominciare*, presentino valori praticamente identici non solo di frequenza token e di frequenza type ma anche nel numero di hapax con cui ciascun verbo ricorre. Tale similarità si riflette anche nella scelta dei nomi che ricorrono con entrambi i verbi con differenze di frequenza di ricorrenza poco significative (cfr. paragrafo successivo sulle preferenze di

selezione). Sulla base di quanto abbiamo detto in precedenza, circa il concetto di sinonimia, potremmo affermare dunque che questi due verbi costituiscono due parole dal significato molto vicino¹³.

Diverso invece il caso di *avviare*, che pur avendo una frequenza token molto elevata, presenta un numero di types e di hapax legomena più basso rispetto agli altri due verbi e di conseguenza un indice di TTR minore che riflette una minore ricchezza lessicale.

Anche nel caso di *terminare/finire+N* i pattern si vanno a collocare su fasce di frequenza distinte. *Finire+N* ha frequenza nettamente maggiore rispetto a *terminare+N* (10502 tokens vs. 3482 tokens). Entrambi i verbi però si caratterizzano per un vocabolario abbastanza ampio e un indice di TTR elevato (*finire+N*, 17,26% mentre *terminare+N*, 21,91%), rispetto ad altri verbi che si collocano nelle medesime fasce di frequenza.

Un caso opposto è invece costituito da *adempiere*, *assolvere* e *ottemperare*. Tutti e tre sono caratterizzati da una frequenza molto bassa nel corpus (*adempiere+comp_a*, 721 tokens; *assolvere+comp_a* 695, tokens; *ottemperare+comp_a*, 441 tokens) e da un vocabolario poco numeroso e semanticamente ristretto. Inoltre, in particolare, *assolvere* e *adempiere*, stabiliscono con determinati nomi legami ben radicati, prototipici (cfr. paragrafo successivo), mentre subiscono un netto calo nella frequenza di ricorrenza degli altri argomenti.

Tuttavia è altrettanto interessante notare come il verbo *ottemperare*, pur ricorrendo con frequenza più bassa rispetto agli altri due, abbia un vocabolario più ampio e un indice di TTR più elevato: *assolvere a*, 10,22%, *adempiere a*, 12,48%, *ottemperare a*, 24,49%. Inoltre, a differenza di *adempiere* e *assolvere*, nel caso di *ottemperare+N*, i valori di frequenza con cui il verbo ricorre con gli argomenti, anche nei primi ranghi, calano meno rapidamente.

Prima di procedere con il lavoro vi è una considerazione importante da fare: dal momento che abbiamo utilizzato un corpus per approssimare l'esperienza linguistica dei parlanti, il modello che se ne ricava è rappresentativo di un certo tipo di testo e di un

¹³ L'idea della similarità come vicinanza semantica è cara agli approcci distribuzionali i quali partono dall'assunto che le parole che tendono a combinarsi con elementi linguistici simili vengono a collocarsi in punti dello spazio semantico più vicini rispetto a quelli occupati da parole che invece si distribuiscono in maniera diversa nel testo. La predisposizione dell'uomo a concettualizzare la similarità come vicinanza all'interno di uno spazio geometrico è spiegata da Sahlgren (2006:18-19) in questo modo: "We are, after all, *embodied* beings, who use our unmediated spatio-temporal knowledge of the world to conceptualize and make sense of abstract concepts. [...] this *similarity-is-proximity* metaphor is so prevalent that it is very difficult to think about similarities, let alone to talk about them, without using it. This also applies to meanings: it is intuitive, if not inevitable, to use the *similarity-is-proximity* metaphor when talking about similarities of meaning. Words with similar meanings are conceptualized as being *near* each other, while words with dissimilar meanings are conceptualized as being *far apart*."

determinato registro, di cui il corpus si fa portavoce, in questo caso quello giornalistico. Le conclusioni che si traggono circa la produttività di un determinato pattern in un corpus non sono totalmente generalizzabili ad altri tipi di linguaggio, pertanto un pattern che risulta produttivo all'interno di un testo potrebbe esserlo in misura differente in un altro.

3.5.2 Differenze nella selezione argomentale.

Un altro aspetto interessante che è emerso dall'analisi dei pattern estratti, riguarda la selezione dei nomi con cui ricorre ciascun verbo all'interno delle coppie di (quasi) sinonimi.

Se la selezione argomentale potesse essere spiegata soltanto in termini di semantica lessicale, coppie di verbi quasi sinonimi dovrebbero, in linea teorica, accettare gli stessi argomenti o comunque nomi appartenenti alle medesime classi semantiche.

Dunque, assumendo pure che non esistano sinonimi perfetti, interscambiabili in ogni contesto *salva veritate*, è abbastanza ragionevole supporre che esistano verbi quasi sinonimi i cui argomenti in linea di principio non dovrebbero appartenere a classi semantiche molto differenti.

In realtà, spesso, all'interno delle coppie di sinonimi è possibile individuare oltre a un insieme, anche consistente, di argomenti condivisi da entrambi i verbi, delle tendenze di selezione argomentale idiosincratiche per ciascuno di essi.

Si pensi alla coppia *rompere* e *infrangere*. I due verbi non presentano differenze sostanziali di significato, entrambi, infatti, indicano l'azione di mandare in frantumi, fare in più pezzi un oggetto fisico o astratto.

Con la stessa naturalezza, infatti, diremo di aver *rotto un equilibrio* o di averlo *infranto*, che un *vetro si è rotto* o *si è infranto*.

Tuttavia, al di là di usi di carattere più idiomatico, in cui la forma si cristallizza per veicolare un significato ben preciso che non è ricostruibile attraverso l'unione dei significati dei suoi componenti (*rompere le scatole* nel senso di “infastidire pesantemente, scocciare” o *rompere il ghiaccio* con il significato di “affrontare per primi una situazione, prendere per primi un'iniziativa, vincere l'imbarazzo”), per ognuno di essi possiamo individuare dei nomi che sono selezionati solo da un verbo e che non ricorrono con l'altro. Per esempio, per parlare della fine di una relazione si dice

di aver *rotto una relazione* e non di aver *infranto una relazione*, allo stesso modo, si parla di *infrangere la legge* e non di *rompere la legge*.

Esistono poi, una serie di nomi che ricorrono con entrambi i verbi ma con pattern di frequenza distinti. I parlanti, infatti, riconosceranno l'espressione *rompere gli schemi* come un modo di dire comune della lingua italiana, ma potrebbero trovare in qualche misura accettabile anche *infrangere gli schemi* che, sebbene più rara, è comunque attestata all'interno del corpus. In questo caso, potremmo ipotizzare che la prima, data la sua elevata frequenza e il suo radicamento nella lingua, sia funta da modello analogico sulla base del quale è stata coniata la seconda.

Allo stesso modo, il corpus riporta l'espressione *infrangere un divieto* con una frequenza relativa¹⁴ molto più elevata (1,595%) rispetto a *rompere un divieto* (0,048%) che comunque è attestato anche se con frequenza molto più bassa.

Vediamo altri casi analoghi:

Figura 13 Differenze di selezione e di frequenza per la coppia *rompere-infrangere*.

nome	rompere	infrangere
Fronte	1,345%	0,106%
Alleanza	1,325%	0,106%
Incantesimo	1,020%	0,053%
Rapporto	2,737%	0,319%
Tregua	0,925%	0,319%
Sogno	0,057%	2,711%
Mito	0,038%	1,489%
Segreto	0,095%	1,276%
Record	0,048%	1,276%
Principio	0,114%	1,116%

Tali differenze di selezione riguardano molti dei verbi quasi sinonimi che abbiamo estratto e in questo paragrafo ci dedicheremo ad analizzarle più nel dettaglio.

Un caso analogo alla coppia *rompere-infrangere* è rappresentato da *riempire-colmare*. Anche in questo caso i due verbi presentano una notevole affinità di significato, entrambi indicano l'azione di far pieno qualcosa con una sostanza o con oggetti anche

¹⁴ Dal momento che, le coppie di sinonimi, hanno rappresentazioni di frequenza differenti all'interno del corpus, per confrontare le frequenze con cui i nomi ricorrevano con ciascuno dei due sinonimi abbiamo lavorato sulle frequenze relative, calcolando il rapporto fra la frequenza assoluta di ciascuna espressione e il numero totale di osservazioni di un pattern.

non materiali (*riempire un vuoto/colmare un vuoto* ma anche *riempire di gentilezze e colmare di gentilezze*).

A differenza di *riempire*, *colmare* è spesso utilizzato all'interno di espressioni idiomatiche come *colmare il sacco* o *colmare la misura* con il significato di "oltrepassare i limiti, arrivare, in cose spiacevoli a un limite intollerabile". Inoltre, così come abbiamo accennato in precedenza, *colmare* stringe con determinati nomi legami molto forti che si traducono in valori elevati di frequenza all'interno del corpus. Tali pattern V+N acquisiscono, in virtù di un forte legame di associazione reciproca, un valore quasi collocazionale e i nomi che accompagnano il verbo, il ruolo di argomenti prototipici: si pensi a espressioni come, *colmare una lacuna*, *colmare un vuoto*.

Possiamo, infatti, notare come i primi ranghi, in un ordine di frequenza discendente, siano occupati da espressioni caratterizzate da frequenze molto elevate come *colmare+vuoto* (616 tokens), *colmare+lacuna* (418 tokens), *colmare+buco* (151 tokens), man mano che si scende, però si nota che le frequenze diminuiscono in modo molto rapido, cosicché basta scendere al rango 24 per trovare un'espressione come *colmare+handicap* con solo 9 occorrenze. Lo stesso andamento non si verifica invece con il verbo *rompere*, per il quale notiamo una diminuzione nella frequenza di ricorrenza dei nomi con il verbo, più lenta.

Per quanto riguarda le differenze di selezione, i due verbi si sovrappongono nella scelta di determinati argomenti: entrambi ricorrono con frequenze elevate con nomi come *vuoto*, *lacuna*, *silenzio*, *spazio*, ecc. Esiste poi un insieme di nomi che ricorrono solo con uno dei due verbi ma non con l'altro, e viceversa (*colmare un ritardo* vs. ??*riempire un ritardo* e *riempire una sala* vs. ??*colmare una sala*) e infine, nomi che ricorrono con frequenze elevate con un verbo, ma basse con l'altro e viceversa (*colmare un divario* e *riempire un divario*, *riempire una pagina* e *colmare una pagina*). A tale riguardo è interessante notare, per esempio, come l'espressione *riempire un bicchiere* sia del tutto naturale mentre *colmare un bicchiere*, per quanto accettabile, sia più marcata; tuttavia, è perfettamente naturale e accettabile dire che *un bicchiere è colmo d'acqua* piuttosto che *è riempito d'acqua*.

Se collochiamo i pattern in un ordine decrescente di frequenza, tali differenze sono percepibili fin dai primi ranghi.

Figura 14 Primi dieci argomenti di *riempire* e *colmare*. Gli argomenti in corsivo e sottolineati, ricorrono con entrambi i verbi con frequenze elevate, i nomi in corsivo ricorrono con entrambi i verbi ma con differenze di frequenza significative.

riempire	frequenza	colmare	frequenza
<i><u>Vuoto</u></i>	509	<i><u>Vuoto</u></i>	616
<i>Pagina</i>	420	<i><u>Lacuna</u></i>	418
<i>Stadio</i>	178	<i><u>Buco</u></i>	151
Bocca	161	<i>Divario</i>	146
<i><u>Spazio</u></i>	159	Deficit	105
<i>Piazza</i>	142	Ritardo	90
Sala	118	<i>Gap</i>	66
Teatro	78	Differenza	64
Strada	78	Distanza	64
Vita	78	Carenza	50

Anche nel caso di *condurre-guidare* notiamo delle differenze interessanti. Entrambi, quando utilizzati in modo transitivo, indicano le azioni di “accompagnare qualcuno facendogli da guida”, “essere a capo di qualcosa”, “dirigere”, “guidare un veicolo”: si dice *guidare un paese* e *condurre un paese*, *guidare una campagna* e *condurre una campagna*, *guidare un’azione* e *condurre un’azione*.

Anche in questo caso, però, oltre a una serie di argomenti che sono comuni a entrambi i verbi, è possibile delineare una serie di tendenze idiosincratiche che portano i verbi a differire nella selezione di determinati nomi.

Per esempio, il corpus attesta una presenza massiccia di *condurre+trasmissione* (frequenza relativa pari a 1,363%), mentre, lo stesso nome, con il verbo *guidare*, è attestato ma con frequenze nettamente più basse (0,075%); allo stesso modo, *guidare+rivolta* ricorre con una frequenza relativa pari a 1,264% laddove *codurre+rivolta* ha scarsa rappresentazione (0,051%).

Esistono poi nomi che sono selezionati da uno solo dei verbi: *guidare un governo* vs. ??*condurre un governo*, *condurre uno studio* vs. ??*guidare uno studio*.

Si veda la tabella riportata sotto per analizzare tendenze analoghe:

Figura 15 Alcune preferenze di selezione dei verbi *guidare-condurre*.

nome	guidare	condurre
classifica	3,506%	0,144%
movimento	0,608%	0,076%

coalizione	0,943%	0,034%
delegazione	2,809%	0,00%
società	1,251%	0,00%
guerra	0,150%	1,803%
esperimento	0,041%	1,337%
sondaggio	0,048%	0,677%
vita	0,00%	5,315%
programma	0,00%	2,124%

Un caso, secondo me interessante, è rappresentato dalla coppia *imparare* e *apprendere*. I due verbi presentano un significato comune: “acquistare cognizione di qualcosa; fare propria e trattenere nella mente una serie di cognizioni attraverso lo studio, l’esercizio, l’esempio altrui”. Oltre a questo, però, *apprendere* presenta un secondo significato che indica il “venire a conoscenza di qualcosa, venire a sapere” che non condivide con *imparare*; fanno capo a questo secondo significato, espressioni come *apprendere una decisione*, *apprendere un risultato*, *apprendere una novità*.

È tuttavia interessante notare come determinati nomi che ricorrono con *apprendere*, utilizzato con questo secondo significato, siano attestati nel corpus, sebbene in misura nettamente inferiore, anche con *imparare*. Per esempio, *apprendere+notizia* si colloca al rango 1 in un ordine decrescente di frequenza, ricorrendo nel corpus con frequenza molto elevata, pari a 691 tokens, con un netto distacco rispetto al pattern di rango 2, *apprendere+lezione* che ricorre 99 volte nel corpus. Tuttavia, il corpus riporta un caso (hapax legomena) in cui anche il verbo *imparare* ricorre con *notizia*:

28) [...] ma non si rendeva conto che era Bill, parlandogli, che riusciva a imparare notizie ed informazioni. [La Repubblica, 2122683]

La frase risulta particolarmente interessante perché attesta un uso analogo anche di *imparare+informazione*.

Esistono altri casi del medesimo tipo, come *apprendere+verità* e *imparare+verità*.

29) Il frassino del cosmo da cui Wotan rimase appeso tre giorni e tre notti per imparare la verità? [La Repubblica, 2247226]

30) Quotidianamente imparava una piccola e disonesta verità da un giornale qualsiasi. [La Repubblica, 15793253]

Anche in questo caso *imparare+verità* ricorre nel corpus con una frequenza relativa pari a 0,160%, molto inferiore rispetto all'attestazione del medesimo nome con *apprendere*, pari a 1,001%.

È difficile però trovare casi in cui il verbo *apprendere*, nel suo significato più prossimo a *imparare*, ricorra in modo esclusivo con nomi che non ricorrono con quest'ultimo. Per farlo, bisogna analizzare fasce di frequenza più basse in cui troviamo attestazioni esclusive di *apprendere* con argomenti come *sintassi* (3 tokens), *principio* (3tokens). Tuttavia si potrebbe pensare che sulla base di esemplari come *imparare la verità* o *imparare informazioni*, il verbo possa essere esteso ad altri argomenti come *novità* o *idea* per il quale al momento, almeno nel corpus preso in analisi, non esistono attestazioni¹⁵.

Non è vero però il contrario: *imparare* infatti ricorre con determinati argomenti in modo esclusivo. Per esempio, il corpus attesta *imparare+lotta* (0,958%) ma non vi sono attestazioni di *apprendere+lotta*; lo stesso si dica per argomenti come *democrazia* (0,607%), *battuta* (0,479%), *confronto* (0,415%). Certi argomenti invece ricorrono con *imparare* con frequenza più elevata rispetto che con *apprendere*: *imparare+parola* (2,077%) vs. *apprendere+parola* (0,316%); *imparare+rispetto* (0,671%) vs. *apprendere+rispetto* (0,053%); *imparare+lavoro* (0,575%) vs. *apprendere+lavoro* (0,053%).

Si veda ora la coppia *alzare-sollevare*. In entrambi i casi i verbi significano “spostare verso l’alto, innalzare, spingere in alto”: con la stessa naturalezza infatti diciamo *sollevare un coperchio* e *alzare un coperchio*, ma anche *sollevare la testa* e *alzare la testa*, *alzare un polverone* e *sollevare un polverone*. *Sollevare* presenta anche una serie di occorrenze figurate in cui assume il significato di *causare, far sorgere, provocare* (*sollevare un dubbio, sollevare una perplessità*) che non sono attestate con *alzare*, se non in misura molto ridotta (*sollevare+protesta* e *alzare+protesta*, entrambi attestati nel corpus anche se con frequenze differenti); mentre il secondo assume anche significato

¹⁵ Un'ulteriore precisazione è opportuna. Nuovi casi d'uso si verificano in prima istanza nella lingua parlata, e solo in un secondo momento, se soggetti a uso particolarmente frequente e se rispecchiano i vincoli grammaticali imposti dalla lingua, trovano attestazione nella lingua scritta. Sarebbe dunque interessante indagare corpora di parlato per verificare se determinate generalizzazioni sono già presenti o lo sono in misura più massiccia nella lingua parlata rispetto a quella scritta.

di “aumentare” (*alzare i prezzi, alzare il livello, alzare la voce*) che non è mai attestato con *sollevare*¹⁶.

Oltre a questi casi, però, esistono determinati argomenti che pur ricorrendo con entrambi i verbi mostrano una netta predilezione per uno soltanto di essi: tale tendenza si traduce in stime differenti di frequenza di attestazione all’interno del corpus. Infine, anche in questo caso, esistono nomi che sono attestati solo con uno dei due verbi ma non con l’altro.

Vediamo alcuni esempi.

Figura 16 Preferenze di selezione per *alzare-sollevare*.

nome	alzare	sollevare
protesta	0,086%	2,241%
vespaio	0,016%	1,104%
peso	0,094%	0,465%
ondata	0,039%	0,639%
telefono	1,586%	0,097%
pugno	0,730%	0,043%
bandiera	3,660%	0,065%
spalle	2,709%	0,043%
polemica	0,00%	5,488%
interrogativo	0,00%	2,067%
palla	0,440%	0,00%
elicottero	0,275%	0,00%

Un altro caso interessante è rappresentato dalla coppia *placare-calmare*: entrambi significano “ridurre a una condizione di calma e tranquillità”. Nonostante la vicinanza di significato, dimostrata da una notevole sovrapposizione nella scelta dei nomi che entrambi selezionano (*animo, acque, tensione, ansia, ira, timore ecc.*) anche in questo caso possiamo notare delle idiosincrasie per ciascun verbo.

Si veda la tabella che riassume tali divergenze:

Figura 17 Differenze di selezione argomentale per la coppia *calmare-placare*.

nome	calmare	placare
------	---------	---------

¹⁶ Questo secondo significato di *alzare*, nel senso di *aumentare*, non è stato preso in considerazione poiché non presenta attestazioni con *sollevare*.

polemica	0,526%	7,745%
fame	0,225%	1,797%
sete	0,676%	2,045%
protesta	0,526%	1,983%
Mercato	1,878%	0,496%
Nervi	1,503%	0,062%
Bollori	1,052%	0,124%
Situazione	1,427%	0,124%
Gente	1,202%	0,124%

Se analizziamo le frequenze con cui i due verbi ricorrono con determinati nomi, notiamo che per esempio, un'espressione come *placare una polemica* ricorre con frequenza molto più elevata rispetto a *calmare una polemica* allo stesso modo in cui *placare la sete* o *placare la fame* hanno frequenza maggiore rispetto a *calmare la fame/sete*.

Al contrario, *calmare i nervi* ha una maggiore attestazione di *placare i nervi* così come l'espressione *calmare i bollori* rispetto a *placare i bollori* nonostante sia rappresentata nel corpus, con bassa frequenza.

Esiste poi una serie di nomi che ricorrono solo con un verbo ma non con l'altro. Il corpus riporta attestate espressioni come *calmare il mare*, *calmare un pensiero*, *calmare un gioco* mentre gli stessi nomi non sono attestati con *placare*, allo stesso modo in cui nomi come *critica*, *clima*, *sospetto*, *dubbio*, ricorrono con *placare* ma non con *calmare*. Vediamo ora i tre verbi, *cominciare*, *iniziare*, *avviare*. Già a livello intuitivo si percepisce che *cominciare-iniziare* presentano un grado di sinonimia maggiore rispetto a *cominciare-avviare* e *avviare-iniziare*. I vocabolari riportano per *cominciare* e *iniziare* definizioni analoghe: “dare inizio, compiere i primi atti di un'azione o di una serie di operazioni; dare principio a qualcosa”, mentre *avviare* oltre al significato di “dare inizio a qualcosa” presenta anche quello di “mettere in via, indirizzare a un luogo determinato” che i primi due non hanno.

Se analizziamo gli argomenti con cui i tre verbi ricorrono notiamo subito che *iniziare* e *cominciare* presentano differenze molto marginali:

Figura 18 Primi dieci nomi, in ordine decrescente di frequenza, selezionati da *cominciare* e *iniziare*. In corsivo i nomi comuni a entrambi nei primi 10 ranghi, sottolineati i verbi che ricorrono con entrambi i verbi nelle prime 20 posizioni.

cominciare	Frequenza relativa	iniziare	Frequenza relativa
<i>Carriera</i>	2,951%	<i>Carriera</i>	3,690%
<i>Lavoro</i>	2,581%	<i>Lavoro</i>	2,782%
<i>Anno</i>	2,210%	<u>Attività</u>	2,587%
<i>Giornata</i>	1,944%	<i>Processo</i>	2,095%
<i>Processo</i>	1,406%	<u>Trattativa</u>	1,734%
<i>Viaggio</i>	1,294%	Sciopero	1,658%
<u>Battaglia</u>	1,252%	<i>Anno</i>	1,471%
<u>Stagione</u>	1,231%	<i>Giornata</i>	1,380%
<u>Vita</u>	1,154%	<i>Viaggio</i>	1,325%
<u>Guerra</u>	1,098%	<u>Esame</u>	1,276%

Nonostante la tabella riporti soltanto i nomi che ricorrono nei primi dieci ranghi, per motivi esemplificativi e per una questione di spazio, già da queste prime posizioni, possiamo avere un'idea di come i due verbi si sovrappongano nella selezione degli argomenti, riportando spesso stime di frequenza per i singoli pattern molto vicine. Inoltre, come avevamo già notato nel paragrafo precedente, questa coppia è caratterizzata da un valore analogo di TTR, che riflette un indice simile di ricchezza e varietà lessicale.

La questione cambia se invece confrontiamo la coppia *cominciare-avviare*¹⁷. I due verbi presentano un insieme consistente di nomi condivisi (*processo, lavoro, attività, esame, trattativa, operazione, stagione*) ma anche tendenze idiosincratiche di selezione argomentale. Vediamone alcune:

Figura 19 Preferenze di selezione *cominciare-avviare*.

nome	cominciare	avviare
sciopero	0,895%	0,051%
giornata	1,944%	0,020%
viaggio	1,294%	0,020%
vita	1,154%	0,026%
partita	1,063%	0,036%

¹⁷ In questo caso non prenderemo in considerazione il significato di *avviare* "mettere in via, indirizzare a un luogo determinato" che non è presente in *cominciare* e che comunque trova una rappresentazione marginale all'interno del corpus.

procedura	0,063%	4,452%
iniziativa	0,035%	1,652%
procedimento	0,007%	1,505%
risanamento	0,049%	0,770%
riflessione	0,091%	0,760%

Lavorando su fasce di frequenza molto più basse, un caso interessante è rappresentato dai verbi *assolvere-adempiere-ottemperare* seguite dal complemento indiretto *comp_a*.

Nel paragrafo precedente avevamo visto come essi siano caratterizzati da un vocabolario poco numeroso e semanticamente ristretto che si traduce in bassi valori di frequenza type e di TTR.

In questa sede ci concentreremo sul confronto fra *adempiere* e *ottemperare* poiché presentano differenze di selezione più interessanti. I due verbi sono riportati, dai vari dizionari come sinonimi anche se presentano una certa divergenza di significato: *ottemperare* infatti indica l' "obbedire, seguendo quanto è prescritto o richiesto" mentre *adempiere*, nella suo utilizzo intransitivo¹⁸ significa "tener fede a qualcosa; compiere, eseguire a pieno".

Se analizziamo gli argomenti con cui ricorrono entrambi i verbi, ci accorgiamo che essi si sovrappongono nella selezione di numerosi argomenti: *obbligo, impegno, dovere, condizione, richiesta, esigenza, necessità, incombenza, volontà* ecc. Ognuno di essi si distingue tuttavia per il fatto di ricorrere frequentemente con argomenti che invece ricorrono raramente con l'altro o di selezionare determinati nomi che invece non sono selezionati dall'altro e viceversa.

Figura 20 Alcuni esempi delle differenze di selezione argomentale fra *adempiere a-ottemperare a*.

nome	Adempiere a	Ottemperare a
compito	11,234%	0,227%
ruolo	2,635%	0,227%
formalità	1,803%	0,454%
funzione	8,599%	0,00%
missione	1,942%	0,00%

¹⁸ Il verbo *adempiere*, in origine transitivo, ha accolto più recentemente un utilizzo intransitivo, accompagnato dalla preposizione *a*. La scelta di analizzare gli argomenti selezionati dalla variante intransitiva deriva dal fatto che quest'ultima ha una maggiore rappresentazione nel corpus che si riflette in un numero di tokens più elevato, quindi è in grado di fornire un campione più numeroso di dati su cui compiere la nostra indagine.

risoluzione	0,416%	4,535%
regola	0,277%	1,814%
legge	0,416%	2,721%
accordo	0,00%	2,494%
invito	0,00%	3,401%

Ingoiare-inghiottire, presentano un significato molto simile nel senso di “buttare giù avidamente, far passare un cibo o una bevanda dalla bocca nella faringe e nell’esofago”. Le differenze di selezione, in questo caso, sono forse più marginali e riconducibili a casi isolati.

Interessante il caso di *ingoiare+rospo* e *inghiottire+rospo*. L’espressione *ingoiare il rospo*, rappresenta un idiomma della lingua italiana con il significato di “costringersi a sopportare qualcosa di molto sgradevole o umiliante perché non si può fare altrimenti, possibilmente senza manifestare il proprio disappunto”, sulla base di tale espressione, deriva probabilmente, per via analogica l’espressione *inghiottire il rospo* che ricorre nel corpus con frequenza più bassa rispetto alla prima ma che ha comunque ampia rappresentazione (14,965% vs. 2,186%) Si vedano i seguenti casi:

31) Non ha potuto far altro che prendere atto delle indicazioni e inghiottire il rospo
(La Repubblica, 355026)

32) Gli israeliani sono stanchi di inghiottire rospi (La Repubblica, 1643036)

Inoltre, probabilmente sempre per analogia con tale espressione, derivano altre espressioni come *ingoiare/inghiottire+sconfitta*, *ingoiare/inghiottire+rabbia*, *ingoiare/inghiottire+umiliazione*, che ricorrono nel corpus ma con frequenza minore.

Anche in questo caso, i verbi si differenziano nella selezione di certi nomi che sono attestati con un verbo e non con l’altro.

Il corpus attesta usi di *inghiottire* con nomi come *auto*, *terra*, *immagine* che invece non sono attestati con *ingoiare* e allo stesso tempo casi contrari, di nomi che ricorrono con *ingoiare* ma non con *inghiottire*, come *ingoiare+delusione*, *ingoiare+emozione*, *ingoiare+idea*.

Infine, ci sono nomi che tendono a ricorrere con maggiore predilezione con uno dei due verbi ma sono attestati anche con l’altro, anche se in misura minore: *inghiottire+strada* (0,830%) ricorre con una frequenza maggiore rispetto a *ingoiare+strada* (0,099%), lo stesso si dica per *inghiottire+uomo* (1,230%) e *ingoiare+uomo* (0,297%),

inghiottire+casa (1,366%) e *ingoiare+casa* (0,396%). Al contrario *ingoiare+sconfitta* ha una frequenza maggiore rispetto a *inghiottire+sconfitta* così come *ingoiare+veleno* rispetto a *inghiottire+veleno*, *ingoiare+amaro* rispetto a *inghiottire+amaro*¹⁹.

Anche la coppia *incorrere-incappare* presenta alcuni casi interessanti. I verbi indicano il “venirsi a trovare in qualcosa di negativo o di spiacevole; cadere in un’insidia; imbattersi in cosa o persona che reca danno o pregiudizio”. A sostegno della quasi sinonimia dei due verbi, il fatto che ricorrono con un numero consistente di argomenti in comune: *sanzione, errore, reato, ira, incidente, infortunio, rigore, accusa, problema, gaffes, difficoltà, guaio, caduta* e molti altri.

Anche in questo caso però ci sono dei nomi che ricorrono con frequenza maggiore con uno dei due verbi rispetto all’altro e nomi che ricorrono solo con uno dei due verbi e non con l’altro e viceversa.

Figura 21 Esempi di differenze di selezione per *incappare-incorrere*.

nome	incappare	incorrere
Multa	0,169%	1,665%
Equivoco	0,253%	1,139%
Violazione	0,253%	1,315%
Provvedimento	0,169%	1,139%
Scomunica	0,169%	1,315%
Legge	1,097%	0,175%
Sconfitta	1,181%	0,088%
Blocco	2,110%	0,088%
Controllo	1,688%	0,351%
Disavventura	1,603%	0,526%

Per quanto riguarda *incappare*, il corpus riporta argomenti come *posto, rete, giornata, inchiesta, vicenda, pattuglia, scandalo, storia* che sono attestati esclusivamente come complementi di *incappare*. Viceversa, i nomi *perdita, omissione, responsabilità,*

¹⁹ Dal momento che le occorrenze, in questo caso, spesso rientravano in fasce di frequenza molto basse, abbiamo utilizzato la funzione *sketch-difference* di *sketch-engine*, selezionando come corpus di riferimento itWak, il corpus di maggiori dimensioni attualmente disponibile, per verificare che le stesse tendenze che abbiamo individuato con Lexit trovassero un secondo riscontro. I risultati, confermano le differenze di frequenza evidenziate in Lexit e ne mostrano di nuove, come il caso di *ingoiare+esca* (149 occorrenze) che in Lexit è attestato solo due volte e che, in entrambi i corpora, non trova attestazione con *inghiottire*.

inesattezza, penalizzazione, delitto, procedimento, sospetto, ricorrono solo con *incorrere*.

Un altro caso interessante è rappresentato da *sradicare-estirpare*. Entrambi indicano l'azione di svellere da terra una pianta con tutte le radici, ma, in senso figurato anche il levar via del tutto, debellare: *estirpare il terrorismo, estirpare la corruzione, sradicare un pregiudizio, sradicare un vizio*.

Lexit attesta usi analoghi di *sradicare* ed *estirpare* con nomi come: *malapianta, pianta, criminalità, terrorismo, mafia, corruzione, pregiudizio, abitudine, idea, memoria*.

Possiamo individuare però delle tendenze idiosincratiche riassumibili nella tabella riportata sotto.

Figura 22 Differenze di selezione per *estirpare-sradicare*.

nome	Frequenza estirpare	Frequenza sradicare
cancro	7,023%	0,613%
bubbone	2,007%	0,204%
violenza	1,672%	0,204%
erbacce	2,341%	0,00%
razza	1,003%	0,00%
albero	0,00%	9,611%
pallone	0,00%	4,090%
avversario	0,00%	1,022%
povertà	0,00%	1,022%
malattia	0,334%	2,454%

Vi è dunque una tendenza di nomi come *cancro, violenza, bubbone, razza* a ricorrere esclusivamente, o con frequenza più elevata con il verbo *estirpare* rispetto a *sradicare*; e una tendenza opposta per altri nomi, fra cui *albero, pallone, avversario, povertà, malattia*.

Anche in questo caso tuttavia, la bassa frequenza di ricorrenza dei due pattern *estirpare+N* e *sradicare+N* non ci permette di verificare in modo chiaro nette tendenze di selezione.

Altri verbi presentano preferenze di selezione più marginali come le coppie *comprare-acquistare* e *terminare-finire*.

Nel primo caso i vocabolari riportano come definizione, per entrambi i verbi, “ottenere la proprietà di un bene, pagandone il relativo prezzo”. *Acquistare* assume però ulteriori

significati, nel senso di “procurarsi, procacciarsi” (*acquistare fama, acquistare esperienza, acquistare forza*), “guadagnare” (*acquistare terreno, acquistare tempo, acquistare altezza*), “aggiungere” (*acquistare merito, acquistare fama, acquistare valore*).

Il confronto ha riguardato la sfera semantica condivisa dai due verbi, nella quale notiamo una netta sovrapposizione nella scelta dei nomi con cui entrambi ricorrono.

Figura 23 Primi dieci argomenti, in ordine di frequenza decrescente per *comprare-acquistare*. In corsivo gli argomenti che sono condivisi da entrambi nelle prime dieci posizioni.

comprare	Frequenza relativa	acquistare	Frequenza relativa
<i>Azione</i>	5,655%	<i>Azione</i>	6,213%
<i>Casa</i>	4,895%	<i>Titolo</i>	4,365%
<i>Titolo</i>	4,008%	<i>Quota</i>	3,127%
<i>Biglietto</i>	2,880%	<i>Biglietto</i>	2,494%
Giornale	1,850%	<i>Diritto</i>	2,382%
Libro	1,767%	<i>Casa</i>	2,224%
Auto	1,519%	Pacchetto	1,766%
<i>Diritto</i>	1,519%	Partecipazione	1,396%
<i>Quota</i>	1,512%	Società	1,349%
Azienda	1,278%	Prodotto	1,344%

I casi, isolati, in cui i verbi si differenziano riguardano soprattutto l'utilizzo di *comprare* accompagnato da nomi che indicano oggetti più piccoli o di valore più basso: *comprare+giornale* (246 tokens) ricorre più spesso che *acquistare+giornale* (55 tokens), *comprare+disco* (111 tokens) ha frequenza maggiore di *acquistare+giornale* (33 tokens), *comprare+pane* (88 tokens) rispetto a *acquistare+pane* (8 tokens). Anche in questi casi, tuttavia, notiamo come le differenze di frequenza, per quanto siano anche consistenti, riguardino sempre fasce alte o medio-alte. Nella sfera opposta, *acquistare* si distacca da *comprare* prevalentemente quando assume significati secondari, in cui non si implica sempre l'ottenere tramite scambio di denaro: *acquistare il potere, acquistare il controllo, acquistare valore*, alcuni di questi argomenti sono attestati anche con *comprare* ma con una diversa accezione del termine.

Anche *terminare* e *finire*, presentano differenze abbastanza marginali. Oltre a un numero insieme di argomenti condivisi (*carriera, lavoro, studio, anno, giornata,*

discorso, periodo, opera, intervento), il corpus riflette alcune leggere idiosincrasie: per esempio, possiamo notare che argomenti come *benzina, soldi, storia* ricorrono con frequenza più elevata con il verbo *finire* piuttosto che con *terminare*. Al contrario, nomi come *operazione, visita, colloquio* sono più frequenti con *terminare* che con *finire*. Ma sono differenze più marginali e sicuramente meno marcate di quelle che abbiamo visto in precedenza.

Infine, ci sono casi in cui all'interno delle coppie di quasi sinonimi uno di essi costituisce una sorta di sottoinsieme dell'altro nella scelta dei propri argomenti.

Questo è il caso di *proibire* rispetto a *vietare*, *paventare* rispetto a *temere*, *impaurire* nei confronti di *spaventare*, *deplorare* rispetto a *criticare*: tutti sono accomunati dal fatto che il primo verbo ricorre nel corpus con frequenza molto più bassa rispetto al secondo.

In altri casi invece i verbi, non presentano nette differenze di frequenza e si sovrappongono nella scelta degli argomenti: *detestare* e *odiare*, *gridare* e *urlare*. Per questo motivo non ho ritenuto opportuno analizzarli nel dettaglio in questa sede.

A partire da queste osservazioni, abbiamo voluto indagare da una parte, fino a che punto un determinato verbo è disposto a essere generalizzato a argomenti con cui non è attestato o è attestato ma con bassa frequenza.

Per farlo, abbiamo deciso di sottoporre determinati casi al giudizio di parlanti nativi italiani per verificare se vi fosse una tendenza maggiore da parte di uno dei due verbi della coppia ad accettare nuovi elementi lessicali o elementi più rari nello slot argomentale.

3.6 Sintesi

In questo capitolo abbiamo descritto gli obiettivi della nostra ricerca a partire dagli studi che in questo ambito sono stati compiuti precedentemente.

Ricollegandoci, in particolare, allo studio compiuto da Zeldes, abbiamo deciso di lavorare su coppie/terne di verbi italiani quasi sinonimi (accettando tale nozione con tutte le dovute cautele) per verificare da un lato, se sia possibile individuare tendenze idiosincratice nella scelta degli argomenti con cui ciascuno di essi ricorre, dall'altro analizzare se, e in quale misura, uno dei due sia più predisposto a essere generalizzato a nuovi utilizzi.

Per l'estrazione dei dati e per l'esplorazione dei loro profili distribuzionali, ci siamo avvalsi di Lexit, un framework computazionale per l'acquisizione automatica e l'esplorazione di informazioni distribuzionali riguardo a nomi, verbi e aggettivi italiani che fa capo al corpus La Repubblica (utilizzato nella presente analisi) e alla sezione italiana di Wikipedia.

Una volta ripuliti e ordinati, i dati sono stati sottoposti ad analisi statistiche: per ogni verbo sono state calcolate la *frequenza token*, la *frequenza type*, il numero di *hapax legomena* e l'indice di *TTR* (type/token ratio) come stima della ricchezza e della varietà lessicale del vocabolario di un verbo.

Per ogni coppia di verbi abbiamo analizzato le preferenze di selezione argomentale e abbiamo visto come, nonostante la quasi sinonimia, spesso i verbi presentino delle differenze nella selezione degli argomenti con cui ricorrono che non possono essere spiegate facendo ricorso alla sola semantica lessicale.

Partendo da questi presupposti ci siamo chiesti quale fosse la capacità e il limite di ciascun verbo a essere generalizzato a nuovi usi o usi più rari.

Abbiamo così deciso di sottoporre coppie di V+N, che nel corpus ricorrevano con frequenze differenti, al giudizio dei parlanti ai quali è stato chiesto di esprimere su una scala Likert da 1 a 7 quanto ritenevano tali espressioni, accettabili nella lingua italiana.

Nel prossimo capitolo, descriveremo la realizzazione dell'esperimento, la creazione degli stimoli e i risultati che abbiamo ottenuto.

4. Realizzazione del test, analisi dei risultati, prospettive future.

In questo capitolo discuterò della realizzazione dell'esperimento, di come sono stati creati gli stimoli e quale è stata la procedura per effettuare il test. Parlerò, inoltre, della metodologia *crowdsourcing* che è stata utilizzata per implementarlo e lanciarlo in rete, delineandone i pregi e alcuni limiti.

Infine discuterò dei risultati che abbiamo ottenuto e delinearò alcuni spunti per eventuali ricerche future.

4.1 La realizzazione del test.

L'obiettivo dell'esperimento psicolinguistico realizzato nella presente indagine era quello di indagare la capacità dei verbi a essere generalizzati a nuovi usi, intendendo con questo, la loro predisposizione ad essere accettati, dai parlanti, quando accompagnati da nomi che non compaiono mai nel corpus nella posizione del loro slot argomentale o vi compaiono ma con bassa frequenza.

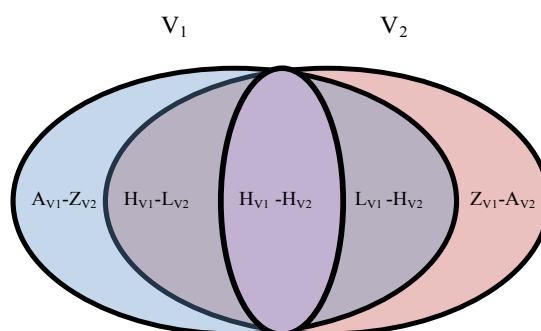
In particolare, l'indagine è stata compiuta su coppie di verbi quasi sinonimi, caratterizzati, per quanto riguarda la selezione degli argomenti, dal ricorrere con un insieme di nomi comuni a entrambi e da preferenze di selezione idiosincratiche per ciascun verbo. Abbiamo voluto verificare il grado di generalizzazione di ciascuno dei due verbi a ricorrere con nomi con i quali, nel corpus, non era attestato o lo era, ma con bassa frequenza, e che invece trovano attestazione con il corrispettivo sinonimo.

4.1.1 Procedura.

Per realizzare gli stimoli dell'esperimento sono state selezionate 10 coppie di verbi quasi sinonimi, intendendo il concetto di sinonimia con tutte le cautele del caso. I criteri alla base della nostra scelta riflettono le differenze di selezione argomentale che, per ciascun verbo, abbiamo delineato nel capitolo precedente. Fra tutti i verbi estratti, abbiamo selezionato quelle coppie di verbi che, pur nella sinonimia di significato presentassero oltre ad un insieme di argomenti condivisi, anche delle differenze interessanti di selezione argomentale nel senso di argomenti che, rispettivamente,

ricorressero o solo con uno dei due verbi ma non con l'altro o con frequenza maggiore con uno rispetto che con l'altro, e viceversa. Là dove erano stati selezionate triple di sinonimi, ne sono stati scelti due sempre nel rispetto dei criteri sopra delineati.

Figura 24 Rappresentazione grafica della sovrapposizione delle coppie di quasi sinonimi nella selezione argomentale.



I verbi che abbiamo selezionato sono i seguenti:

- Ingoiare-inghiottire
- Incorrere in-incappare in
- Imparare-apprendere
- Cominciare-avviare
- Calmare-placare
- Adempiere-ottemperare
- Guidare-condurre
- Rompere-infrangere
- Riempire-colmare
- Alzare-sollevar

Per ogni coppia di verbi sono stati scelti 25 nomi sulla base di 5 condizioni:

- 5 nomi che ricorressero con frequenza elevata con entrambi i verbi (H-H)²⁰.
- 5 nomi che ricorressero con frequenza elevata con un verbo e bassa frequenza con l'altro (H-L), e viceversa (L-H).

²⁰ La dicitura H, indica la frequenza elevata (high), L, frequenza bassa (low), A indica che l'espressione è attestata, Z indica che non ci sono attestazioni nel corpus di quel nome insieme a quel determinato verbo.

- 5 nomi che fossero attestati con un verbo ma non con l'altro (A-Z), e viceversa (Z-A).

Al fine di verificare che le differenze di frequenza individuate nel corpus fossero effettivamente significative, sui dati delle tre condizioni H-H, H-L, L-H, è stato eseguito il test di Student, quando la distribuzione dei dati risultava normale, in caso contrario, è stato utilizzato il test di Wilcoxon.

Nella prima condizione, ci siamo accertati che il test non individuasse una differenza statisticamente significativa (p-value >0,05) dimostrando appunto che non c'era una diversa distribuzione fra i due gruppi; mentre per la seconda e la terza condizione, abbiamo appurato che ci fosse una differenza statisticamente significativa (p-value <0,05) fra i nomi che ricadevano nella condizione H e quelli che ricadevano nella condizione L per ciascun verbo.

Una volta raccolti i vari pattern V+N, secondo queste caratteristiche, abbiamo creato le espressioni-stimolo utilizzate nel task linguistico.

Le espressioni sono state realizzate lasciando il verbo all'infinito seguito dal nome, preceduto dall'articolo: *incorrere in un rischio - incappare in un rischio*.

Gli stessi nomi sono stati presentati per entrambi i verbi della coppia.

Per la scelta dei nomi, il primo criterio che abbiamo seguito è stato quello della frequenza: i nomi dovevano ricorrere con differenze di frequenza nette, a parte nella condizione (H-H), fra un verbo e l'altro. Nel caso in cui, vi fossero nomi che ricorrevano con pattern di frequenza simili, abbiamo selezionato quello, a nostro giudizio, più idoneo a comparire nel test, tenendo conto, soprattutto, del fatto che i soggetti non avevano l'ausilio del contesto per valutare il grado di accettabilità di un'espressione. Inoltre sono stati scartati i forestierismi o nomi che necessitassero di un'ulteriore specificazione per essere compresi.

Per quanto riguarda invece la scelta dell'articolo, abbiamo preferito utilizzare, quando possibile, l'articolo indeterminativo; in ogni caso, gli stessi nomi, quando presentati con un verbo e con l'altro sono stati accompagnati dal medesimo articolo per evitare che quest'ultimo influenzasse il giudizio dei soggetti nel fornire il giudizio di accettabilità.

Sulla base di questi criteri, sono stati realizzati 500 stimoli, la cui lista completa, è disponibile in appendice.

Qui sotto riporto un esempio a titolo esemplificativo.

Figura 25 Esempio di espressioni utilizzate nel test. A fianco è riportata l'informazione circa la frequenza con cui l'espressione era attestata nel corpus La Repubblica.

espressione	frequenza
Rompere un equilibrio	H
Infrangere un equilibrio	H
Rompere un'alleanza	H
Infrangere un'alleanza	L
Rompere un divieto	L
Infrangere un divieto	H
Rompere un contratto	A
Infrangere un contratto	Z
Rompere una legge	Z
Infrangere una legge	A

Gli stimoli così creati, sono stati presentati, uno alla volta e in ordine casuale, ad un gruppo di soggetti madrelingua italiani ed è stato chiesto loro, per ogni espressione di fornire un giudizio di accettabilità sulla base di una scala Likert da 1 a 7. Un punteggio uguale a 1 significava che l'espressione era ritenuta dai parlanti innaturale, mentre un'espressione pari a 7 corrispondeva a piena naturalezza e accettabilità.

Nelle istruzioni, sono stati presentati ai soggetti due esempi:

- a) Leggere un libro (massima accettabilità, punteggio 6 o 7)
- b) Leggere una sedia (minima accettabilità, punteggio 1 o 2)

Prima di lanciare l'esperimento, è stato compiuto un *pilot* per verificare che il design dell'esperimento fosse ben fatto e soprattutto che i soggetti avessero chiara comprensione delle istruzioni e, di conseguenza, che le risposte avessero senso in relazione alla domanda posta nel task. Per compierlo sono state selezionate 2 coppie di verbi, nella fattispecie *inghiottire-ingoiare*, *incappare in-incorrere in*, e sono stati raccolti 10 giudizi per ogni coppia V+N.

Dai risultati ottenuti, abbiamo avuto modo di verificare che il task era stato compreso e che le risposte fornite dai soggetti rispecchiavano la loro effettiva comprensione delle istruzioni.

Sia l'esperimento che il pilot, sono stati realizzati con metodologia *crowdsourcing*, un metodo innovativo, che permette, tramite una piattaforma, di lanciare in rete il proprio

task, raccogliendo giudizi grazie al contributo di partecipanti esterni, in tempi rapidi e con costi contenuti.

4.1.2 Il metodo del crowdsourcing.

Per condurre l'esperimento, nella presentazione degli stimoli e nella raccolta delle risposte, è stato utilizzato Crowdfunder (<https://www.crowdfunder.com>), una piattaforma online di crowdsourcing, realizzata nel 2007 da Lukas Biewald e Chris Van Pelt a San Francisco, in California, che permette agli utenti di pulire, etichettare e arricchire i propri dati grazie al contributo di milioni di partecipanti.

La particolarità della piattaforma risiede nel fatto che i dati non sono raccolti attraverso esperimenti di laboratorio ma sul web, grazie al contributo di partecipanti esterni, motivo per cui consente di raccogliere numerosi giudizi in tempi rapidi.

Questa nuova metodologia ha portato numerosi vantaggi, a livello tempistico ed economico nella realizzazione di test di vario tipo. Non è tuttavia esente da alcuni limiti. In particolare, il fatto di non poter supervisionare i soggetti, durante l'esecuzione del task, come avverrebbe in un tradizionale esperimento laboratoriale, porta inevitabilmente alcuni svantaggi.

Il limite principale consiste nel non avere alcun tipo di controllo e feedback sui propri soggetti e sulle loro risposte. Queste ultime, possono differire moltissimo in relazione all'età, all'estrazione sociale, al livello di istruzione, alla provenienza geografica dei partecipanti, fattori questi che non possono essere controllati e che però possono influenzare notevolmente le risposte fornite dai parlanti. Inoltre, dovendo creare un task che sia fruibile da tutti, spesso si rischia di incorrere in eccessive semplificazioni dello stesso che possono condizionare i risultati ottenuti. Non è, infine, possibile monitorare i partecipanti durante l'esecuzione del task, e dunque verificare che questi ultimi lo compiano, nella sua interezza, con la massima attenzione e serietà: negli esperimenti compiuti in laboratorio, chi compie l'esperimento ha la possibilità di interagire con i partecipanti, assicurarsi che le istruzioni siano state ben comprese, individuare possibili cali di attenzione e infine accertarsi, che il livello di conoscenza di una lingua sia adeguato al task da compiere. Venendo a mancare questo tipo di incontro-confronto, certe anomalie nelle risposte fornite dai soggetti possono essere più difficili da interpretare.

Numerosi però sono anche i vantaggi. Tale metodologia consente infatti di raccogliere un grande quantitativo di dati in tempi relativamente circoscritti e con costi contenuti. Per verificare la validità, la qualità e l'attendibilità dei risultati raccolti con questo metodo, rispetto ai metodi tradizionali di laboratorio, sono stati compiuti numerosi studi in campo psicolinguistico. In particolare, è interessante l'articolo scritto da Munro et al. (2010), dal titolo "Crowdsourcing and language studies: the new generation of linguistic data" il quale propone un compendio di lavori e progetti nei quali è stata adottata la metodologia del crowdsourcing e i cui risultati sembrano dimostrare non solo che la loro attendibilità è del tutto comparabile a quella del metodo tradizionale in laboratorio, ma anche che, in alcuni casi, la qualità dei dati ottenuti è perfino migliore. Tali studi spaziano in vari ambiti delle discipline linguistiche, dalla semantica alla psicolinguistica, e nello specifico hanno riguardato la trasparenza semantica, la segmentazione di una traccia audio di linguaggio parlato, la predicibilità contestuale, i giudizi di accettabilità; essi inoltre, forniscono supporto alle evidenze apportate da studi condotti su corpora, e sembrano dimostrare che l'analisi delle onde cerebrali (ERP) possa essere migliorata attraverso un'analisi crowdsourcing degli stimoli sperimentali. Munro propone, anche alcuni suggerimenti per venire meno agli svantaggi derivati dal mancato controllo dei soggetti durante l'esecuzione del task. Per esempio, consiglia di inserire all'interno del task alcune domande volte a verificare il grado di attenzione dei partecipanti e il loro livello di competenza linguistica per garantire un miglior livello di attendibilità dei risultati.

In conclusione, possiamo dire che la metodologia del crowdsourcing costituisce uno strumento utile e innovativo per riprodurre, su larga scala, numerosi esperimenti tradizionalmente compiuti in laboratorio e raccogliere un grande quantitativo di dati, in tempi brevi e con costi contenuti. La validità dei dati ottenuti consente di generare risultati empirici che rispondano a domande e ipotesi riguardo a numerosi fenomeni linguistici.

Tuttavia, qualora non si ritenga opportuno affidare la propria ricerca esclusivamente a tale metodologia, è sempre possibile, verificare la validità dei risultati ottenuti riproponendo, in un secondo momento, il medesimo task ad un gruppo di soggetti in un contesto tradizionale di laboratorio.

4.1.3 I partecipanti.

Al test hanno partecipato 15 soggetti, madrelingua italiani, caratteristica quest'ultima che è stato chiesto esplicitamente di rispettare, nelle istruzioni del test; i partecipanti provengono da varie regioni italiane.

Il numero di partecipanti non è elevato, ma abbiamo ritenuto che fosse comunque sufficiente a delineare delle tendenze circa il fenomeno preso in analisi.

4.2 I risultati.

I dati raccolti sono stati ordinati e suddivisi per coppie di verbi, in modo da poter paragonare differenze e similarità di comportamento all'interno delle coppie di sinonimi. Per ciascun pattern V+N sono stati raccolti 15 giudizi per un totale di 75 giudizi per categoria, 375 giudizi per verbo e 750 per coppia di sinonimi.

Per ogni espressione è stata calcolata la media dei giudizi di accettabilità forniti dai soggetti e la varianza, come indicatore della variabilità dell'insieme di dati; quest'ultima è risultata un indicatore particolarmente interessante per l'analisi di determinate espressioni: ci sono casi, infatti, in cui la stima della varianza dei giudizi raccolti presenta valori elevati indicando dunque che il gruppo di soggetti a cui è stata sottoposta l'espressione, non è stato concorde nel fornire il proprio giudizio di accettabilità.

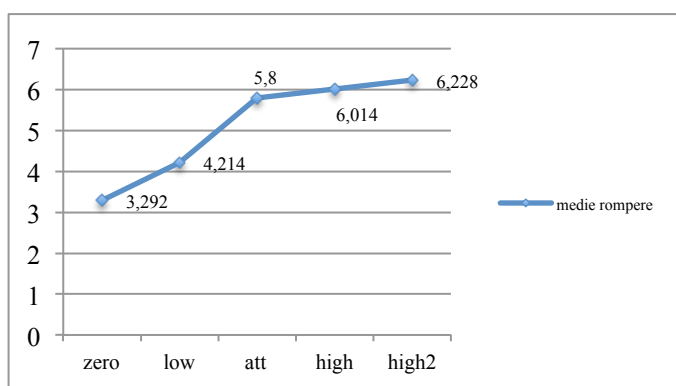
Il calcolo della media è stato effettuato anche sul totale dei giudizi per ciascuna categoria (H_2 , H_1 , L, A, Z)²¹. Possiamo vedere come, in un grafico a linee, se disponiamo in ordine crescente sul piano delle ascisse le categorie prese in analisi (Z, L, A, H_1 , H_2)²² e, su quello delle ordinate, i valori dei giudizi di accettabilità (da 1 a 7), ne emerge un andamento crescente dei giudizi a partire dalle espressioni non attestate (che presentano i valori più bassi) fino alle espressioni attestate con frequenze alte.

²¹ Con la dicitura H_2 si indica quella categoria di nomi che trovano elevata attestazione con entrambi i verbi, mentre con H_1 , la categoria di nomi che sono attestati con frequenza elevata solo con uno dei due membri della coppia e che invece sono attestati con frequenze basse con l'altro.

²² La categoria A, come abbiamo visto sopra, identifica un insieme di nomi che sono attestati con solo uno dei due verbi della coppia di quasi-sinonimi, andandosi a incrociare con la categoria Z. Nel realizzare il grafico, tale categoria è stata messa fra la categoria L e quella H poiché racchiude casi che nel corpus erano attestati con stime di frequenza variabili ma tendenzialmente alte.

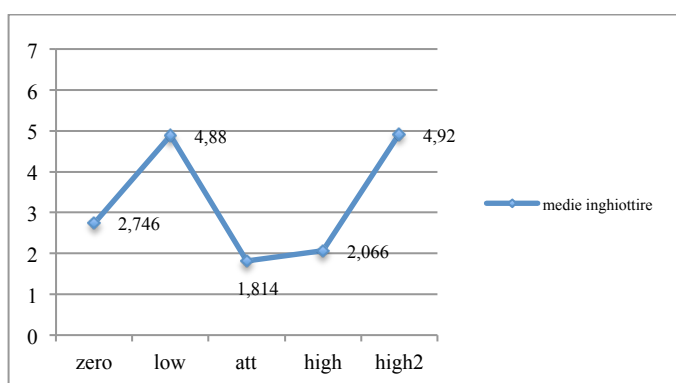
Come possiamo vedere dal grafico riportato in fig. 3, in cui è preso come esempio il verbo *rompere*, i giudizi di accettabilità forniti dai parlanti, per le varie categorie, incrementano dalla categoria Z, che identifica un insieme di nomi che nel corpus non trovano mai attestazione con il verbo *rompere*, fino alla categoria H, che ha ottenuto, prevedibilmente, le stime più alte di accettabilità.

Figura 26 Rapporto fra l'aumentare dei giudizi di accettabilità rispetto alle categorie di attestazione delineate nel corpus La Repubblica per i verbi; nella fattispecie, a titolo esemplificativo riportiamo il caso di *rompere*.



Tale andamento crescente si verifica per le medie dei giudizi di accettabilità di tutti i verbi con un'eccezione, costituita dai giudizi di accettabilità espressi per il verbo *inghiottire*.

Figura 27 Medie dei giudizi di accettabilità rispetto alle categorie di attestazione nel corpus, La Repubblica, per il verbo *inghiottire*.



Come possiamo vedere, i giudizi di accettabilità seguono un andamento non lineare rispetto alle categorie di attestazione individuate nel corpus.

Nello specifico, possiamo vedere come i soggetti abbiano attribuito giudizi mediamente più alti alle espressioni appartenenti alle categorie Z e L, rispetto a quelle appartenenti

alle categorie A (nomi attestati nel corpus con *inghiottire*) e H (nomi che nel corpus erano attestati con *inghiottire* con frequenza elevata).

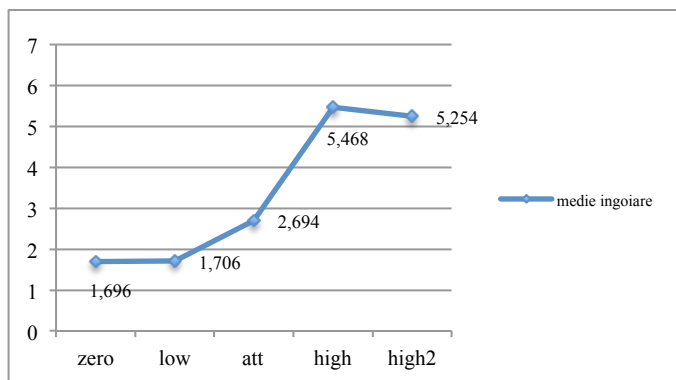
Tale anomalia trova una spiegazione se analizziamo i nomi che abbiamo selezionato per le due categorie. Per quanto riguarda le categorie A e H, sono stati scelti, per creare gli stimoli, prevalentemente nomi concreti, *casa, strada, uomo, persona, auto, terra, città, immagine*, ecc.; al contrario, nelle categorie L e Z, troviamo, per la maggior parte nomi astratti, *sconfitta, umiliazione, idea, delusione, emozione*, ecc. Probabilmente, tale discrepanza fra il giudizio di accettabilità fornito dai soggetti e la frequenza con cui i nomi trovavano attestazione nel corpus, si nasconde proprio in una distinzione fra concetti concreti e astratti. È possibile, in questo caso particolare, che, nella mancanza di un adeguato contesto linguistico/pragmatico, i soggetti abbiano ritenuto più accettabile l'idea di inghiottire concetti astratti, come una delusione, una sconfitta (anche sulla base di un'estensione analogica rispetto all'espressione *inghiottire un rospo*), piuttosto che oggetti concreti che solitamente non sono ingeriti da un essere vivente. Si deve altresì considerare che *ingoiare* e *inghiottire*, in quanto verbi di ingestione, prevedono come soggetto prototipico, nel ruolo di AGENTE, un essere animato; il fatto che nel task, siano state presentate espressioni prive della specifica del soggetto che compie l'azione, ha probabilmente portato i soggetti a compiere automaticamente tale associazione (verbo di ingestione-soggetto animato) e a percepire un'anomalia quando il verbo è stato presentato con oggetti, in questo senso, marcati, non prototipici. Sarebbe interessante, dunque, riproporre gli stessi pattern V+N, in presenza di un adeguato contesto linguistico, all'interno di espressioni come *la frana inghiottì la casa* vs. *la frana ingoiò la casa*, per verificare se i risultati presentano un esito diverso.

È interessante notare, inoltre, come la media dei giudizi raccolti per ciascuna espressione presenti un indice elevato di varianza, ad indicare una certa variabilità nei punteggi forniti dai soggetti e quindi una mancanza di accordo nell'esprimere i giudizi. Questo è il caso di espressioni come *inghiottire una casa* (varianza pari a 1,928), *inghiottire le risorse* (varianza pari a 2,007), *inghiottire un'azienda* (1,776).

Il fatto che, *ingoiare*, non presenti lo stesso andamento anomalo è dovuto al fatto che, in questo caso, la situazione, è ribaltata: nel corpus infatti, i nomi astratti ricorrevano con il verbo già di per sé con frequenza elevata, mentre i nomi concreti erano attestati in misura minore o non lo erano affatto. Ne risulta un andamento più lineare in cui per i

nomi appartenenti alle categorie Z e L sono stati raccolti giudizi di accettabilità più bassi, mentre per le categorie A e H, giudizi mediamente più alti.

Figura 28 Medie dei giudizi di accettabilità rispetto alle categorie di attestazione nel corpus, La Repubblica, per il verbo *ingoiare*.



Come notiamo nel grafico, vi è un incremento nella predisposizione dei parlanti ad accettare le espressioni man mano che si va dalla categoria Z costituita da pattern V+N che nel corpus non trovavano attestazione, a H, espressioni che, al contrario, vi vantavano una presenza massiccia.

Ricapitolando, abbiamo detto che per ogni coppia di verbi sono stati selezionati 25 nomi suddivisi per 5 condizioni:

- 1) H-H, nomi che ricorressero nel corpus con frequenza elevata con entrambi i verbi.
- 2) H-L, nomi che ricorressero con frequenza elevata con un verbo ma bassa con quell'altro.
- 3) L-H, condizione speculare alla precedente.
- 4) A-Z, nomi che fossero attestati solo con uno dei due verbi e non con l'altro
- 5) Z-A, condizione speculare alla precedente.

Per ciascuna condizione, gli stessi nomi sono stati presentati con entrambi i verbi, per un totale di 50 stimoli a coppia.

Analizziamo ora, i risultati che sono scaturiti dall'applicazione del test di Student per confrontare i giudizi di accettabilità raccolti, per ogni coppia di sinonimi, all'interno di

ogni condizione, per ciascuna categoria a confronto (H-H; H-L; L-H; A-Z; Z-A). Analizzeremo ogni coppia separatamente.

Inghiottire – ingoiare.

Dall'analisi del t-test, compiuto sui giudizi di accettabilità raccolti per le varie categorie a confronto, per la coppia *inghiottire-ingoiare* emerge quanto segue.

Nella condizione $H_{(\text{inghiottire})} - H_{(\text{ingoiare})}$ ²³, non vi è una differenza statisticamente significativa nella distribuzione dei dati (p-value 0,3381), questo significa che, per i nomi selezionati, i parlanti non presentano una maggiore predilezione ad usare uno dei due verbi piuttosto che un altro, e riflette altresì, la situazione che abbiamo individuato nel corpus, ovvero la presenza di un insieme di nomi che effettivamente possono ricorrere con entrambi i membri della coppia, senza distinzione alcuna.

Fra le varie espressioni prese in analisi, è interessante il confronto che emerge fra *ingoiare il rospo* e *inghiottire il rospo*. La prima costituisce un modo di dire, un idioma della lingua italiana, con il significato di “costringersi a sopportare qualcosa di spiacevole o umiliante perché non si può fare altrimenti”; la seconda, che trova comunque larga attestazione nel corpus, è stata creata probabilmente tramite un procedimento analogico sulla base della prima. Se compiano il t-test sui giudizi espressi per entrambe le espressioni, il p-value pari a 0,2824, mostra che i soggetti non sembrano percepire una differenza quando la stessa espressione è presentata con il verbo prototipico (*ingoiare*) o con il sinonimo. In questo senso, potremmo parlare di un processo di generalizzazione verificatosi nella direzione opposta rispetto a quella che stiamo analizzando in questo studio. Il fenomeno di estensione, in questo caso, non ha riguardato lo slot argomentale, ma quello verbale.

Il caso è particolarmente interessante per due motivi: da una parte, perché perfettamente in linea con quanto afferma Zeldes (2012) riguardo al fatto che, le strutture argomentali, rispetto ai pattern morfologici, presentano più di uno slot argomentale e la scelta di quali di essi considerare determinanti nell'analisi del fenomeni della produttività, dipende dal tipo di analisi che si vuole compiere; dall'altra perché riflette un'idea della produttività come un continuum che va da schemi più aperti, generali e regolari a estensioni analogiche compiute su un unico esemplare.

²³ Per non stare a ripetere, ogni volta, a cosa corrispondono le varie condizioni, e per facilitare la lettura dei risultati, da ora in avanti verrà utilizzata tale dicitura: in lettere maiuscole la categoria che riporta la frequenza di ricorrenza dei nomi con i verbi e in pedice, il verbo a cui fanno capo i nomi per quella categoria.

Nella condizione $H_{(\text{inghiottire})} - L_{(\text{ingoiare})}$, non vi è differenza significativa nei giudizi espressi dai soggetti quando gli stessi nomi, sono stati presentati con un verbo o con l'altro (p-value=0,1348). Questo dato riflette l'anomalia di cui abbiamo parlato sopra circa la maggiore predisposizione dei soggetti, in mancanza di un contesto linguistico e/o pragmatico opportuno, ad accettare con maggiore facilità, un verbo di ingestione quando accompagnato da un nome astratto piuttosto che da un nome concreto che non appartiene alla classe degli oggetti prototipici che possono essere ingeriti, indipendentemente dalla frequenza con cui i nomi sono attestati effettivamente nel corpus.

Nella condizione opposta, $L_{(\text{inghiottire})} - H_{(\text{ingoiare})}$, quando cioè i nomi che ricorrevano nel corpus con frequenza elevata con *ingoiare* e bassa con *inghiottire*, sono stati presentati con entrambi i verbi, il t-test mostra una differenza al limite della significatività fra i due gruppi (p-value 0,0408). Il valore del p-value vicino allo 0, sta ad indicare che i parlanti non sembrano percepire un differenza netta. Se analizziamo i giudizi espressi per le singole espressioni a confronto notiamo che ci sono casi in cui, al contrario una differenza non è percepita.

Per esempio, nel corpus, avevamo individuato una differenza di frequenza interessante fra i pattern *ingoiare+sconfitta* (attestato con frequenza elevata) e *inghiottire+sconfitta* (bassa frequenza). Il confronto fra i giudizi di accettabilità espressi dai parlanti, per le espressioni *ingoiare una sconfitta* e *inghiottire una sconfitta*, non sembra riflettere tale disparità (p-value= 0,4606); stessa cosa si dica per le espressioni *ingoiare un veleno* e *inghiottire un veleno* (p-value=0,5995), *ingoiare amaro* e *inghiottire amaro* (p-value=0,2364).

Al contrario, emerge una leggera differenza nei giudizi espressi per le espressioni *ingoiare un'umiliazione* vs. *inghiottire un'umiliazione* (p-value= 0,0355).

Nella condizione $Z_{(\text{inghiottire})} - A_{(\text{ingoiare})}$, si verifica la situazione opposta: i risultati non mostrano una differenza significativa (p-value 0,8589) fra i giudizi di accettabilità espressi quando gli stessi nomi, che nel corpus erano attestati con *ingoiare* ma non con *inghiottire*, sono stati presentati come argomenti per entrambi i verbi. Questo dato è importante perché indica che *inghiottire* può essere esteso agli usi che sono fatti di *ingoiare*, nonostante, di tali usi non vi sia attestazione nel corpus. È interessante notare, inoltre, come la media dei giudizi che sono stati espressi per le espressioni che ricadevano nella categoria $Z_{(\text{inghiottire})}$ è di poco più alta (2,746) rispetto alla media dei giudizi forniti per la categoria $A_{(\text{ingoiare})}$ pari a 2,694.

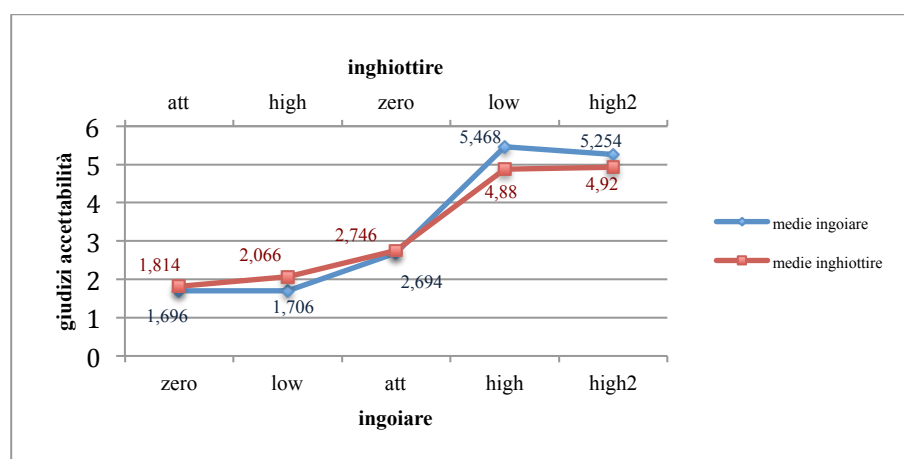
Per la condizione opposta $A_{(\text{inghiottire})} - Z_{(\text{ingoiare})}$, il t-test riporta un p-value pari a 0,5433. Anche in questo caso, un ruolo importante è stato svolto dalla presenza, come stimoli, di nomi concreti che non costituiscono oggetti prototipicamente ingeribili e dalla concomitante mancanza di un contesto linguistico adeguato; l'anomalia percepita dai parlanti trova riflesso nella media dei giudizi espressi per entrambe le categorie: $A_{(\text{inghiottire})}$, 1,814; $Z_{(\text{ingoiare})}$, 1,696. È quindi difficile trarre conclusioni certe. Possiamo comunque dire che, in mancanza di contesto linguistico, i parlanti non sembrano percepire una distinzione quando gli stessi nomi, che indicano oggetti concreti non commestibili, sono presentati come argomenti di entrambi i verbi, nonostante nel corpus, fossero attestati solo con *inghiottire*.

In generale, in virtù di quanto abbiamo detto sopra, non è possibile per questa coppia di verbi, trarre delle conclusioni precise circa la maggiore predisposizione di uno dei due membri della coppia ad essere esteso a nuovi usi o usi più rari.

Come già accennato, sarebbe comunque interessante, ripetere il task fornendo un maggiore contesto linguistico per avere un prospettiva più chiara circa tendenza dei due verbi ad essere generalizzati.

La situazione, può comunque essere riassunta attraverso un grafico in cui sul piano delle ordinate sono collocati i giudizi di accettabilità su una scala da 1 a 7, e, sui due piani delle ascisse, rispettivamente le categorie di attestazione del verbo *inghiottire* nel corpus e quelle speculari di *ingoiare*.

Figura 29 Confronto fra le medie dei giudizi di accettabilità espressi per *ingoiare* e per *inghiottire*.



Dal grafico possiamo vedere che le medie dei giudizi di accettabilità, forniti per ciascuna categoria, si sovrappongono in quasi tutti i nodi del grafico, con eccezione

delle categorie $L_{(\text{inghiottire})} - H_{(\text{ingoiare})}$, in cui leggermente divergono e che corrispondono alla condizione per cui abbiamo trovato, attraverso il t-test, una differenza significativa.

Incappare – incorrere.

Il t-test, eseguito per ciascuna categoria a confronto, riporta i seguenti risultati.

Nella condizione $H_{(\text{incappare})} - H_{(\text{incorrere})}$, il test non riporta una differenza statisticamente significativa per i due gruppi (p-value 0,1933), a sostegno del fatto che i due verbi presentano un insieme di nomi per i quali sono perfettamente interscambiabili, anche secondo il giudizio dei soggetti.

Nella condizione $H_{(\text{incappare})} - L_{(\text{incorrere})}$, il test riporta una differenza al limite della significatività (p-value 0,0429). I soggetti sembrano prediligere espressioni formate da *incappare*, accompagnato dai suoi argomenti più frequenti, piuttosto che estendere il loro uso a *incorrere*, nonostante il corpus ne riporti attestazioni. Tuttavia, il fatto che il valore del p-value, pur rientrando nel range della significatività statistica, sia comunque molto vicino alla soglia dello 0,05, indica che, più che una netta predilezione, si tratti di una tendenza.

Una situazione analoga si verifica per la condizione $L_{(\text{incappare})} - H_{(\text{incorrere})}$; anche in questo caso, il test riporta una distribuzione statisticamente diversa per i due gruppi (p-value 0,0436), con un valore del p-value vicino alla soglia della significatività. Ancora una volta, si può parlare di una tendenza dei parlanti a prediligere l'uso di *incorrere* quando accompagnato dai suoi argomenti prototipici piuttosto che di una generalizzazione di *incappare* agli stessi (seppure, anche in questo caso, nel corpus siano attestati casi in cui il verbo ricorre con questi stessi argomenti).

Interessanti i risultati che emergono dall'applicazione del test rispettivamente nelle condizioni $A_{(\text{incappare})} - Z_{(\text{incorrere})}$ e $Z_{(\text{incappare})} - A_{(\text{incorrere})}$.

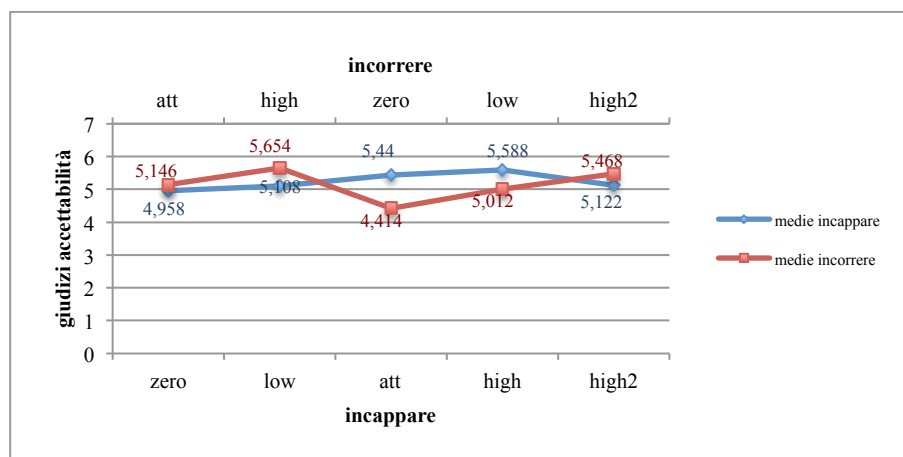
Nel primo caso, il test riporta una differenza statisticamente significativa (p-value= 0,0002) fra i giudizi di accettabilità raccolti relativamente alle espressioni costituite da *incappare*, accompagnato da nomi con cui era effettivamente attestato nel corpus (*incappare in un'inchiesta*, *incappare in uno scandalo*, *incappare in una pattuglia*) e espressioni create con i medesimi nomi presentati come argomenti di *incorrere* (di cui il corpus non riportava attestazioni): *incorrere in un'inchiesta*, *incorrere in uno scandalo*, *incorrere in una pattuglia*. Questo sembra indicare una mancata predisposizione di *incorrere* ad essere generalizzato a nomi con cui non trova mai attestazione, ma che sono attestati con il verbo sinonimico. Non è vera la situazione contraria.

Possiamo vedere, infatti, che nella condizione opposta ($Z_{(incappare)} - A_{(incorrere)}$), la differenza non è statisticamente significativa (0,5036), riflettendo dunque il fatto che i soggetti accettano la generalizzazione degli usi di *incorrere* a *incappare*, anche quando di questi usi il corpus non riporta attestazione. In alcuni casi, sono stati raccolti giudizi mediamente più alti per espressioni costituite dal verbo *incappare* piuttosto che per quelle formate da *incorrere*: *incappare in responsabilità giuridiche*, ha ottenuto una media di giudizi pari a 5,93, pur non essendo mai attestato nel corpus, mentre *incorrere in responsabilità giuridiche*, ha ottenuto una media di 5,4, pur essendo attestata; stessa cosa si dica per *incappare in un'omissione* (media di 4,8) e *incorrere in un'omissione* (media di 4,6).

Il risultato è particolarmente interessante perché in linea con i dati ricavati dall'analisi sulle frequenze dei pattern V+N, che abbiamo compiuto preliminarmente al task, sui dati estratti dal corpus: avevamo notato come il verbo *incappare*, presentasse un indice di TTR maggiore (41,77%) rispetto a *incorrere* (26,99%) e dunque una maggiore ricchezza e varietà lessicale nel vocabolario, dato che ci aveva portato a ipotizzare una maggiore predisposizione di *incappare* a ricorrere con nuovo materiale lessicale nello slot argomentale.

La media dei giudizi di accettabilità, rispetto alle categorie prese in analisi per i due verbi, può essere rappresentata dal grafico riportato in fig.7. Possiamo notare appunto come le linee che rappresentano la variazione delle medie dei giudizi di accettabilità in rapporto alle categorie prese in analisi divergano là dove il t-test ha riscontrato una differenza significativa fra i due gruppi ($H_{(incappare)} - L_{(incorrere)}$, $L_{(incappare)} - H_{(incorrere)}$, $A_{(incappare)} - Z_{(incorrere)}$), e come invece vi sia una sovrapposizione per le condizione $Z_{(incappare)} - A_{(incorrere)}$ e $H_{(incappare)} - H_{(incorrere)}$.

Figura 30 Confronto fra le medie dei giudizi di accettabilità espressi per *incorrere* e *incappare*.



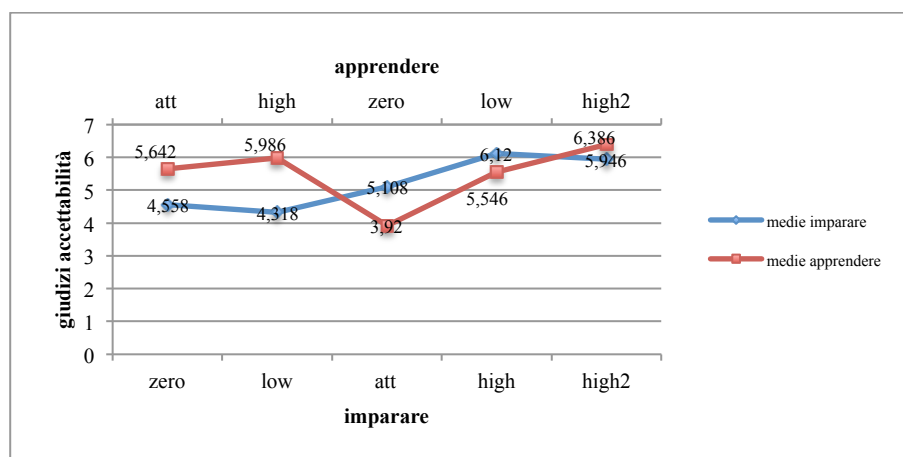
Imparare – apprendere.

Nella condizione $H_{(imparare)} - H_{(apprendere)}$ il t-test non mostra una differenza statisticamente significativa ($p\text{-value} = 0,0691$)²⁴, riflettendo l'effettiva sinonimia dei due verbi nel selezionare un insieme di nomi per i quali sono interscambiabili.

Per tutte le altre condizioni, troviamo differenze significative: $H_{(imparare)} - L_{(apprendere)}$, p-value pari a 0,0267; $L_{(imparare)} - H_{(apprendere)}$, p-value pari a 6,4274E-09; $A_{(imparare)} - Z_{(apprendere)}$, p-value uguale a 0,0002; $Z_{(imparare)} - A_{(apprendere)}$, p-value 0,0004.

Vediamo la situazione rappresentata nel grafico:

Figura 31 Confronto medie dei giudizi di accettabilità espressi per *imparare* e *apprendere*.



Come possiamo evincere chiaramente dal grafico, a parte una certa similarità nei giudizi espressi nella condizione in cui i due verbi sono stati presentati ai soggetti accompagnati

²⁴ Anche se il valore è vicino alla soglia di significatività ($p\text{-value} = 0,05$).

da nomi che avevano elevata attestazione con entrambi, per le altre categorie non vi è la tendenza, da parte di nessuno dei due membri, ad essere esteso ai nomi attestati esclusivamente con il rispettivo sinonimo, o che, con quest'ultimo presentavano maggiore attestazione.

Questo risultato, può essere spiegato, in parte, con il fatto che, il verbo *apprendere*, presenta due significati: quello di “acquistare cognizione di qualcosa”, che condivide con *imparare* e quello di “venire a conoscenza”, non condiviso con il sinonimo. Uno degli obiettivi della nostra indagine, era verificare se i parlanti fossero portati ad accettare *imparare* accompagnato da nomi attestati con il verbo *apprendere*, inteso nel secondo significato delineato. Tale curiosità scaturiva dall'aver osservato, nel corpus, attestazioni di casi rari, talvolta hapax legomena, del tipo *imparare+notizia* e *imparare+informazione*, *imparare+verità*, che danno adito all'ipotesi che si possa verificare una generalizzazione in questa direzione. Abbiamo così pensato di presentare queste espressioni ai soggetti per verificarne il grado di accettabilità.

I risultati mostrano che non vi è una generalizzazione in tal senso. Tuttavia, è interessante notare come la media dei giudizi di accettabilità per le singole espressioni spesso non abbia valori bassi. La media dei giudizi espressi per *imparare un particolare* è pari a 4,93, per *imparare un dettaglio* e *imparare un'informazione* è per entrambi di 4,73; in questi ultimi due casi, l'indice di varianza è abbastanza elevato, 1,843, nel primo caso e 2,015, nel secondo, a indicare una certa variabilità nell'insieme dei dati raccolti, conseguente a un mancato accordo da parte dei soggetti nell'esprimere la propria valutazione. Se compiamo un t-test per verificare la distribuzione dei giudizi di accettabilità sulle singole espressioni, emergono dei risultati interessanti: per quanto riguarda le espressioni *imparare un'informazione* e *apprendere un'informazione*, il test, compiuto sui giudizi di accettabilità restituisce un p-value pari a 0,0129, ai limiti della significatività, mentre per *imparare un dettaglio* e *apprendere un dettaglio* un p-value pari allo 0,051; infine, se applichiamo il medesimo test sui giudizi espressi per *imparare un particolare* e *apprendere un particolare* risulta un p-value pari a 0,273, che riflette una mancanza di diversità nella distribuzione dei giudizi per le due espressioni.

Al contrario, il confronto fra i giudizi espressi per le espressioni *imparare una nozione* e *apprendere una nozione* fa emergere una differenza nettamente significativa (p-value 9,44329E-05), così come per le espressioni *imparare la verità* e *apprendere la verità* (p-value = 0,0016).

Possiamo spiegare l'esito di questo risultato facendo ricorso i dati ricavati dal corpus. Già nel capitolo precedente avevamo notato come *notizia* costituisca, per *apprendere*, l'argomento di rango 1, e come con esso stringa un legame molto forte, prototipico, in virtù della frequenza con cui il pattern ricorre nella lingua, e di riflesso, anche nel corpus (691 volte); avevamo anche sottolineato come la frequenza con cui ricorrono gli altri nomi, subisca un calo repentino a partire dall'argomento di rango 2 (*lezione*), che ricorre 99 volte.

Il radicamento di tale espressione, unito al fatto che essa fa capo al significato di *apprendere*, non condiviso con *imparare*, porta inevitabilmente i parlanti ad accettare in misura minore un'espressione in cui l'argomento prototipico di *apprendere*, *notizia*, ricorre con *imparare*.

Al contrario, quando ai parlanti sono presentate espressioni costituite da *imparare* accompagnato da argomenti frequenti di *apprendere* ma non così radicati, questi, sembrano mostrare una maggiore predisposizione all'estensione.

Avevamo altresì ipotizzato, che sulla base di esemplari come *imparare+verità*, *imparare+informazione*, lo stesso verbo potesse essere esteso anche a nomi con i quali nel corpus non presentava attestazioni ma che invece erano sempre attestati con *apprendere* (condizione $Z_{(imparare)} - A_{(apprendere)}$). Con questo intento, sono stati selezionati i pattern: *imparare+idea* e *imparare+novità*.

Se confrontiamo i giudizi espressi per le singole espressioni *imparare un'idea* rispetto a *apprendere un'idea*, dal t-test scaturisce un p-value pari a 0,0449, al limite della soglia di significatività; al contrario per le espressioni *imparare una novità* e *apprendere una novità*, la differenza è nettamente significativa (p-value= 0,0006); anche in questo caso si può fornire una spiegazione in termini di prototipicità: *novità* costituisce sicuramente un oggetto più prototipico di *apprendere* di quanto non lo sia *idea*²⁵.

Sulla base dei dati riportati, a mio giudizio, si può ipotizzare che un certo fenomeno di generalizzazione incipiente si stia effettivamente verificando, che sta portando *imparare* a accettare nuovi elementi lessicali con cui prima non era accettato e ad una conseguente estensione di significato.

L'applicazione del test sui giudizi di accettabilità espressi per le condizioni $H_{(imparare)} - L_{(apprendere)}$ e $A_{(imparare)}$ e $Z_{(apprendere)}$, restituisce i seguenti risultati. Per quanto riguarda la prima condizione, il p-value è al limite della soglia di significatività (0,0267). Se

²⁵ A sostegno di tale ipotesi, le stime di frequenza ricavate dal corpus: *apprendere+idea* ricorre nel corpus con una frequenza relativa pari allo 0,369%; *apprendere+novità*, con una frequenza relativa pari a 0,948%.

analizziamo i giudizi espressi per le singole espressioni, messe a confronto, vediamo che ci sono casi in cui i parlanti non hanno percepito una differenza fra l'utilizzo di un verbo piuttosto che un altro: il t-test applicato ai giudizi raccolti per *imparare una parola* rispetto a *apprendere una parola*, restituisce un p-value pari a 0,3238, stessa cosa per i giudizi raccolti nei confronti delle espressioni *imparare il rispetto* e *apprendere il rispetto* (p-value= 0,1257), *imparare un lavoro* e *apprendere un lavoro* (p-value= 0,1729); questo significa che certi nomi molto frequenti di *imparare*, che erano attestati anche con *apprendere*, ma in misura minore, sono generalizzabili al secondo verbo.

Per la condizione A-Z, il t-test mostra, al contrario, una differenza nettamente significativa (p-value = 0,0002). I nomi che ricorrono esclusivamente con *imparare* non sembrano accettabili, secondo il giudizio espresso dai parlanti, nello slot argomentale di *apprendere*.

Nel complesso, i risultati emersi dall'applicazione del t-test, sembrano mostrare che i due verbi si comportino in modo complessivamente divergente; pur mostrando, in alcuni casi isolati un fenomeno di generalizzazione incipiente.

Questo andamento trova, in parte, una spiegazione nella discrepanza di significato che contraddistingue i due verbi, nel senso delineato sopra; sarebbe interessante riproporre il test selezionando nomi che fanno capo allo stesso significato per entrambi i verbi, per verificare se vi è una tendenza maggiore da parte di uno dei due ad essere esteso agli argomenti selezionati dall'altro.

È altresì vero che, anche quando sono confrontati sullo stesso significato (nelle condizioni $A_{(imparare)} - Z_{(apprendere)}$ e $H_{(imparare)} - L_{(apprendere)}$) la distinzione permane, lasciandoci presupporre che i parlanti percepiscano una certa distinzione nell'utilizzo di un verbo o dell'altro.

Cominciare-avviare.

Nel capitolo precedente abbiamo visto come *cominciare* e *avviare*, siano spesso riportati nei vocabolari come verbi sinonimici nel significato di “dare inizio a qualcosa”. Analizziamo ora le tendenze di generalizzazione dei due verbi e verificiamo se i soggetti sono più predisposti a generalizzare gli usi di uno dei due membri della coppia a quell'altro, quando questi sono attestati nel corpus, per entrambi i verbi ma con pattern di frequenza differenti.

Anche in questo caso, abbiamo applicato il t-test per confrontare i giudizi di accettabilità espressi dai soggetti nelle varie condizioni: H – H, H – L, L – H, A – Z, Z – A.

Nella condizione $H_{(\text{cominciare})} - H_{(\text{avviare})}$, il test mostra una differenza non statisticamente significativa fra i giudizi espressi per le espressioni appartenenti ad una e l'altra categoria, a dimostrare il riconoscimento da parte dei soggetti di un'effettiva sinonimia fra i due verbi, i quali ricorrono con una serie di nomi con i quali sono interscambiabili.

Nella condizione $H_{(\text{cominciare})} - L_{(\text{avviare})}$, il test riporta una differenza significativa (p-value = 9,25218E-11). Ciò significa che i parlanti percepiscono una differenza di accettabilità quando i nomi che nel corpus ricorrevano con frequenza elevata con *cominciare* gli sono proposti come argomenti di *avviare*, nonostante gli stessi nomi siano attestati, sebbene in misura minore anche con quest'ultimo.

Nella condizione speculare, $L_{(\text{cominciare})} - H_{(\text{avviare})}$, il test mostra una differenza significativa (p-value = 0,0262) con un valore del p, al limite della soglia di significatività. Se però applichiamo il t-test sui giudizi espressi per le singole espressioni, ci sono casi in cui il test non riporta una differenza significativa: *cominciare un progetto* e *avviare un progetto*, (p-value=0,4265). Nel caso di *cominciare una riflessione* e *avviare una riflessione*, non solo non riscontriamo una differenza significativa (p-value 0,1207) ma, in media, l'accettabilità dell'espressione *cominciare una riflessione* risulta più alta (5,2) rispetto ad *avviare una riflessione* (4,07), nonostante quest'ultima sia attestata con maggiore frequenza nel corpus.

Tali risultati sembrano mostrare comunque una tendenza di *cominciare* a essere esteso agli argomenti di *avviare*.

Particolarmente interessanti i risultati emersi dall'applicazione del test alle condizioni $Z_{(\text{cominciare})} - A_{(\text{avviare})}$ e $A_{(\text{cominciare})} - Z_{(\text{avviare})}$.

Nel primo caso, il test non riporta una differenza significativa (p-value =0,9019). I soggetti sembrano propensi a generalizzare il verbo *cominciare* anche ai nomi con i quali nel corpus non ha attestazioni, ma che avevano attestazione con il sinonimo: *cominciare un ripensamento*, *cominciare una razionalizzazione*, *cominciare un riassetto*, *cominciare una creazione*. Alcune espressioni hanno perfino raggiunto mediamente giudizi di accettabilità più alti rispetto a quando gli stessi nomi sono stati presentati con *avviare*. Questo è il caso di *cominciare una creazione* (5,4) rispetto ad *avviare una creazione* (4,6) e *cominciare un riassetto* (4,6) rispetto a *avviare un riassetto* (4,13).

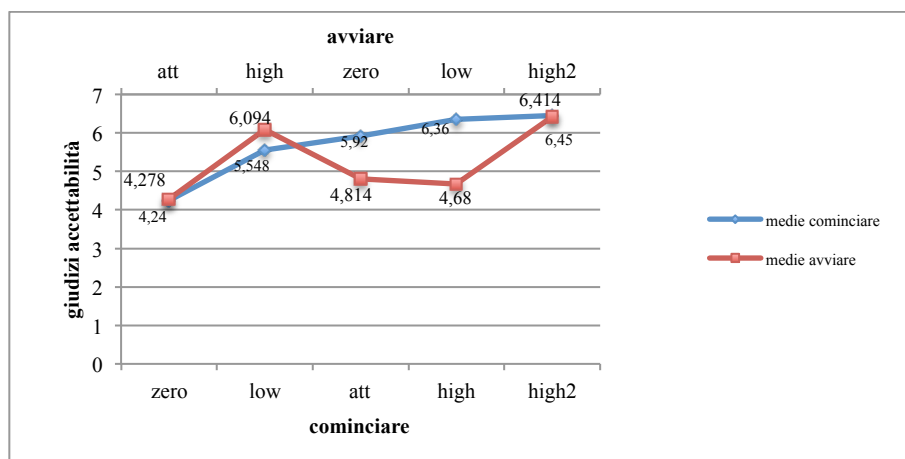
Per la condizione speculare $A_{(\text{cominciare})} - Z_{(\text{avviare})}$, il t- test restituisce una differenza nettamente significativa ($5,74865E-06$). Anche questo dato è importante perché riflette, al contrario della condizione precedente, una mancata generalizzazione degli usi di *cominciare* ad *avviare*. I soggetti, non sembrano portati a estendere i nomi che nel corpus erano attestati esclusivamente con *cominciare*, ad *avviare*.

Alla luce di quanto abbiamo detto finora, i dati ricavati dall'esperimento sembrano riflettere le differenze di TTR delineate nel capitolo precedente. Avevamo visto, infatti, come il pattern *avviare*+N, pur presentano nel corpus un numero totale di occorrenze (19607 tokens) più elevato, rispetto a *cominciare*+N (14299 tokens), presentasse un numero di types e di hapax legomena più basso, riflesso di un vocabolario più ripetitivo; da ciò scaturiva una stima di TTR più bassa (*avviare*+N, 7,43% vs. *cominciare*+N, 10,97%). Avevamo dunque ipotizzato una minore predisposizione da parte del verbo ad accettare nuovo materiale lessicale nella posizione dello slot argomentale, ipotesi che sembra trovare riscontro nei risultati ottenuti attraverso il task.

Questi risultati, sono in linea con quanto afferma Zeldes (2012) circa il rapporto fra il radicamento delle costruzioni nella lingua e la loro produttività. Egli sostiene, che il fatto che una costruzione sia molto comune, all'interno di una lingua, e quindi particolarmente radicata, non implica che sia predisposta ad accettare nuovi elementi lessicali nei suoi slot. Esistono pattern, che pur avendo un numero di ricorrenze elevato, presentano comunque un vocabolario molto ripetitivo, e al contrario, costruzioni rare che invece sono istanziate da un elevato numero di elementi lessicali.

Nel grafico riportato sotto possiamo riassumere i risultati che abbiamo appena descritto.

Figura 32 Confronto medie dei giudizi di accettabilità espressi per *cominciare* e *avviare*.



Calmare-placare.

Nella condizione $H_{(\text{calmare})} - H_{(\text{placare})}$, il t-test compiuto sui giudizi di accettabilità assegnati per le espressioni appartenenti alle due categorie a confronto, non mostra una differenza statisticamente significativa (p-value= 0,3859).

Nella condizione $H_{(\text{calmare})} - L_{(\text{placare})}$, il t-test sui giudizi di accettabilità riporta un p-value pari a 0,0443, ai limiti della significatività.

Se andiamo a confrontare però, i giudizi raccolti per le singole espressioni a confronto, notiamo che, ad eccezione delle espressioni *calmare i bollori* e *placare i bollori*, per le quali si riscontra una differenza significativa nei giudizi ottenuti dai soggetti (p-value = 0,0238), per le altre, non si verifica la stessa situazione: *calmare un mercato* vs. *placare un mercato*, *calmare i nervi* vs. *placare i nervi*, *calmare una situazione* vs. *placare una situazione*, *calmare la gente* vs. *placare la gente* non mostrano differenze statisticamente significative.

È necessario sottolineare che, *calmare i bollori* costituisce un'espressione idiomatica della lingua italiana; in quanto tale, è caratterizzata da un maggior grado di fissità sintattica e semantica; nei confronti di tali espressioni, i parlanti assumono un atteggiamento più conservatore e cauto soprattutto nel generalizzarle a nuovi verbi; tuttavia il fatto che il valore del p-value sia vicino allo 0,05, lascia intuire che essa non sia percepita come un'espressione idiomatica completamente rigida, ma presenti comunque un certo grado di flessibilità.

Inoltre, il fatto che l'espressione *placare i bollori*, abbia ottenuto complessivamente un giudizio di accettabilità pari a 6, lascia intuire che sia considerata dai parlanti un modo di dire pienamente accettabile.

Nella condizione contraria, $L_{(\text{calmare})} - H_{(\text{placare})}$, il test mostra una differenza non significativa, seppure sulla soglia della significatività (p-value= 0,05815).

Una differenza molto più netta si ha nel confronto fra le condizioni $A_{(\text{calmare})} - Z_{(\text{placare})}$ e quella opposta $Z_{(\text{calmare})} - A_{(\text{placare})}$.

Nel primo caso, il t-test compiuto sui giudizi di accettabilità per le due categorie messe a confronto, restituisce un p-value pari a 0,4184; nel secondo, al contrario, mostra una differenza statisticamente significativa (p-value = 0,0015).

Alla luce di questi dati, possiamo inferire che, i due verbi mostrano un discreto livello di sinonimia, questo è dimostrato dalla mancanza di una differenza statistica fra i giudizi espressi nella condizione $H - H$ e dalla tendenza dei soggetti a generalizzare gli usi di un verbo a quell'altro nelle condizioni $H - L$ e $L - H$ nonostante nel primo caso sia

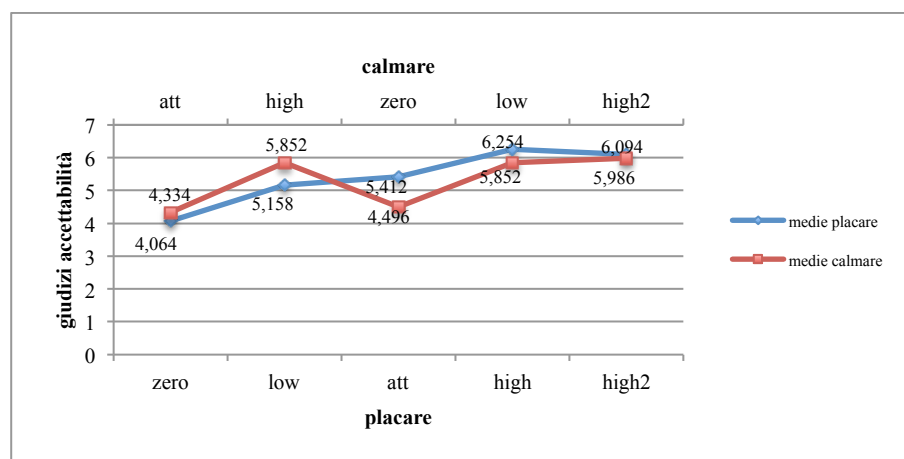
riscontrata una differenza significativa e nel secondo no, i valori del p, in entrambi i casi sono vicini alla soglia della significatività (p-value 0,05).

Per quanto riguarda i nuovi usi, vi è tuttavia una tendenza maggiore di *placare* a essere esteso a nomi con cui nel corpus non è mai attestato ma che trovano attestazione con il sinonimo, *calmare*, dimostrata dalla mancanza di una differenza significativa per la condizione $A_{(calmare)} - Z_{(placare)}$. Al contrario, *calmare*, non sembra essere generalizzabile a nuovi elementi lessicali, attestati esclusivamente con *placare*²⁶.

È comunque necessario sottolineare, come

Possiamo riassumere i risultati ottenuti attraverso il grafico riportato sotto:

Figura 33 Confronto medie dei giudizi di accettabilità espressi per *calmare* e *placare*.



I risultati sono in linea con le tendenze riscontrate nel corpus: confrontando la stima di TTR dei due pattern, emergeva che *placare*+N presentava un valore di poco più elevato rispetto a *calmare*+N (*placare*+N, 25,03% vs. *calmare*+N, 23,06%), riflesso di un vocabolario più ricco e vario, e più predisposto a essere esteso a nuovi elementi lessicali.

Adempiere-ottemperare.

Il t-test compiuto sui giudizi di accettabilità per le categorie a confronto mostra i seguenti risultati.

²⁶ Da considerare il fatto che le espressioni che ricadevano nella condizione A, per *calmare*, hanno raggiunto in media, giudizi di accettabilità bassi (4,334) inferiori anche alla condizione Z.

Nella condizione $H_{(adempiere)} - H_{(ottemperare)}$, non emerge una differenza significativa (p-value=0,6603), a sostegno del fatto che i due verbi presentano un insieme di nomi con cui sono interscambiabili.

Nella condizione $H_{(adempiere)} - L_{(ottemperare)}$, il t-test sui giudizi di accettabilità, restituisce un p-value pari a 0,0014. I soggetti non sembrano propensi a estendere gli argomenti prototipici di *adempiere*, a *ottemperare*, nonostante il corpus presenti attestazioni di questi ultimi anche con il secondo verbo, seppur in misura minore.

Nella condizione opposta, $L_{(adempiere)} - H_{(ottemperare)}$, si verifica la situazione contraria. Il test non riporta una differenza statisticamente significativa (p-value= 0,4464). *Adempiere*, è accettato con gli stessi argomenti di *ottemperare*, nonostante, con il primo, questi fossero attestati con bassa frequenza, mentre per il secondo, costituissero nomi che ricorrevano con frequenza elevata.

Questa situazione è speculare a quella vista sopra e implica un fenomeno di generalizzazione del verbo *adempiere* agli usi di *ottemperare*.

Per quanto riguarda le altre due condizioni, $A_{(adempiere)} - Z_{(ottemperare)}$ e $Z_{(adempiere)} - A_{(ottemperare)}$, il t-test non mostra differenze significative nei giudizi espressi per le condizioni a confronto: $A_{(adempiere)} - Z_{(ottemperare)}$, p-value= 0,2225; $Z_{(adempiere)} - A_{(ottemperare)}$, p-value= 0,3557.

Tale esito mostra una tendenza da parte di entrambi i verbi ad essere generalizzati a nuovi elementi lessicali che, nel corpus, sono attestati esclusivamente con il rispettivo sinonimo.

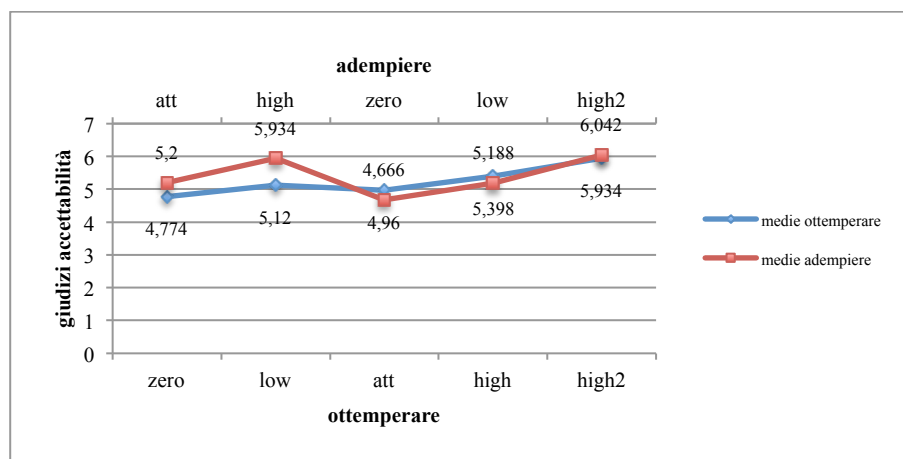
Il fatto che, ad eccezione della condizione $H_{(adempiere)} - L_{(ottemperare)}$ in cui vi è una differenza significativa, per le altre, il t-test non registri differenze, è indicativo del fatto che i parlanti percepiscono i due verbi come effettivamente sinonimici ed interscambiabili in quasi tutti i contesti.

Inoltre, i risultati sembrano mostrare che i parlanti non percepiscano, fra i due verbi, quella differenza di significato che è attestata in alcuni vocabolari, nel senso di *adempiere* come “eseguire a pieno, compiere” e *ottemperare* come “obbedire”, fatto questo che è dimostrato dalla mancanza di differenza fra i giudizi espressi per espressioni come *adempiere a un accordo* (mai attestato nel corpus) vs. *ottemperare a un accordo* (attestato nel corpus), p-value= 0,1708, oppure *adempiere a una mansione* (attestato) e *ottemperare a una mansione* (non attestato), p-value= 0,1638.

La generalizzazione degli usi di un verbo a quelli dell'altro anche là dove questi stessi usi non trovavano attestazione nel corpus, è intuibile attraverso il grafico riportato sotto,

nel quale possiamo vedere come le medie dei giudizi di accettabilità, per le varie categorie a confronto, ad eccezione della condizione ($H_{(adempiere)} - L_{(ottemperare)}$), si sovrappongono nei punti nevralgici del grafico.

Figura 34 Confronto medie dei giudizi di accettabilità espressi per *adempiere* e *ottemperare*.



Guidare-condurre.

Il t-test compiuto sui giudizi di accettabilità per le varie categorie a confronto mostra, nella condizione $H_{(guidare)} - H_{(condurre)}$, una differenza significativa ($p\text{-value} = 0,0394$), anche se al limite della soglia di significatività ($p\text{-value} 0,05$). In generale, per gli stessi nomi, *condurre* ha ricevuto giudizi di accettabilità più alti, nonostante essi fossero attestati, nel corpus, con frequenze elevate con entrambi i verbi. Una possibile spiegazione è che i parlanti non percepiscano una completa sinonimia e interscambiabilità fra i due verbi, anche quando accompagnati da nomi comuni ad entrambi.

Per quanto riguarda le condizioni $H_{(guidare)} - L_{(condurre)}$ e $L_{(guidare)} - H_{(condurre)}$, in entrambi i casi, il t-test riporta differenze nettamente significative: $p\text{-value} = 3,69828E-05$, nel primo caso e $p\text{-value} = 2,19164E-09$, nel secondo. La distinzione nella frequenza con cui i nomi ricorrono con i due verbi, nel corpus, è percepita dai parlanti, che esprimono giudizi differenti quando gli stessi nomi gli sono presentati per una o l'altra condizione, in relazione alla stima della frequenza di ricorrenza nel corpus.

I nomi che nel corpus ricorrevano con frequenza elevata con *condurre* non sembrano essere accettati nella stessa misura come argomenti di *guidare*, nonostante il corpus ne riporti attestazioni, lo stesso vale per la situazione speculare.

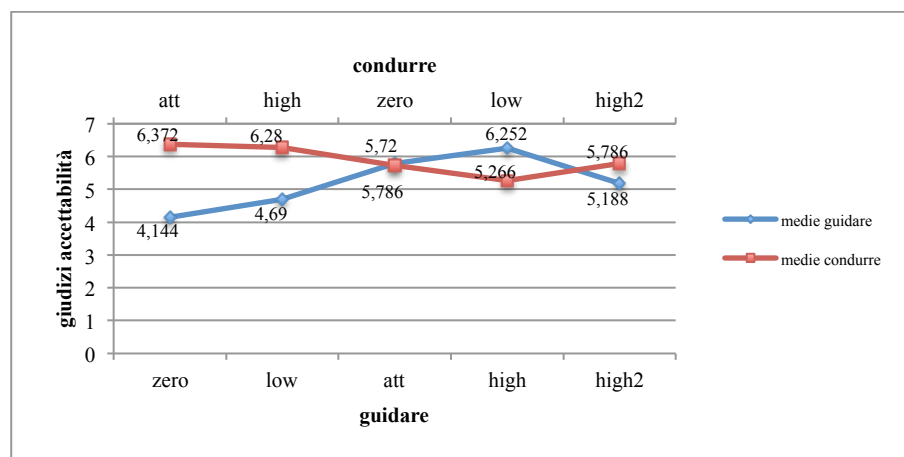
Per quanto riguarda la categoria $A_{(\text{guidare})} - Z_{(\text{condurre})}$, l'applicazione del test non fa emergere una differenza significativa, $p\text{-value} = 0,7693$. Il verbo *condurre* sembra dunque generalizzabile a quei nomi con i quali non è riportata attestazione nel corpus ma che sono, al contrario, attestati con *guidare*.

Nella situazione opposta, $Z_{(\text{guidare})} - A_{(\text{condurre})}$, il test riscontra una differenza statistica netta, $p\text{-value} = 1,38788\text{E-}14$. In opposizione con quanto si è verificato sopra, i soggetti non sono portati a generalizzare i nomi che ricorrono con *condurre* a *guidare*.

È probabile, dato l'esito dei risultati, che i parlanti non percepiscano un elevato grado di sinonimia fra i due verbi e che percepiscano una distinzione nella selezione dei nomi con cui ciascuno di essi può ricorrere.

Dal grafico, possiamo vedere il fenomeno di generalizzazione che avviene nella condizione $A_{(\text{guidare})} - Z_{(\text{condurre})}$ e divergenza nei giudizi espressi per le altre condizioni a confronto, compresa quella H-H.

Figura 35 Confronto medie dei giudizi di accettabilità espressi per *guidare* e *condurre*.



Rompere-infrangere.

Dal t-test compiuto sui giudizi di accettabilità per le varie categorie a confronto emerge quanto segue.

Nella condizione $H_{(\text{rompere})} - H_{(\text{infrangere})}$, nessuna differenza significativa, $p\text{-value} = 0,4128$, a sostegno della sinonimia dei due verbi e del fatto che esiste un gruppo di nomi con cui entrambi possono ricorrere indistintamente senza variare il grado di accettabilità dell'espressione che li contiene.

Per tutte le altre condizioni, $H_{(\text{rompere})} - L_{(\text{infrangere})}$ ($p\text{-value} = 0,0016$), $L_{(\text{rompere})} - H_{(\text{infrangere})}$ ($p\text{-value} = 5,61618\text{E-}11$), $A_{(\text{rompere})} - Z_{(\text{infrangere})}$ ($p\text{-value} = 8,65518\text{E-}06$) e

$Z_{(\text{rompere})} - A_{(\text{infrangere})}$ (p-value= 1,40279E-19), il test riscontra sempre differenze statisticamente significative.

I risultati quindi fanno emergere che i due verbi hanno effettivamente comportamenti di selezione differenti e che nessuno dei due sembra generalizzabile agli elementi lessicali con i quali è esperito raramente o non è esperito affatto ma che sono invece argomenti molto frequenti o comunque attestati col sinonimo. In particolare non sembra esserci da parte di nessuno dei due la tendenza a espandersi nel dominio semantico dell'altro; ognuno, al contrario, rimane strettamente legato ai nomi con i quali è attestato.

Anche in questo caso probabilmente ciò che vincola il fenomeno della generalizzazione è il radicamento dei nomi selezionati con i rispettivi verbi che costituisce un freno al fenomeno della produttività, come abbiamo visto precedentemente con Zeldes (2012).

Sia *rompere* che *infrangere*, stringono con gli argomenti più frequenti, legami particolarmente forti. *Rompere* ricorre all'interno di molti usi idiomatici: *rompere il ghiaccio*, *rompere le scatole*, *rompere gli indugi*, *rompere gli argini*). *Infrangere*, invece tende a stringere con i propri argomenti più frequenti, legami molto forti, di carattere collocazionale: si pensi a espressioni come *infrangere le regole*, *infrangere la legge*, *infrangere un sogno*.

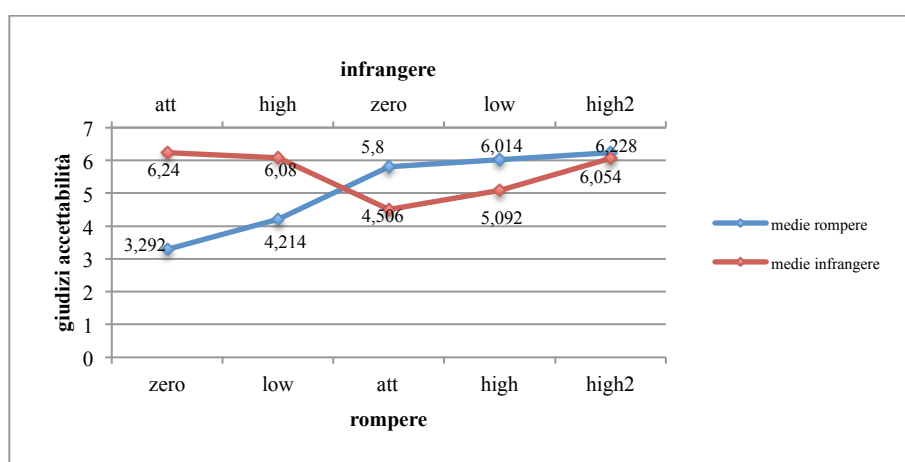
In linea con quanto sostenuto da Zeldes (2012), la probabilità che un pattern ricorra con nuovo materiale lessicale, dipende dalla forza del radicamento del pattern stesso in rapporto a quello dei suoi types. Gli argomenti più frequenti, lessicalizzati e, in particolare, le collocazioni, sviluppano con i verbi che li selezionano, connessioni forti. Come conseguenza, accade che l'intero pattern sia associato dai parlanti, principalmente a questi argomenti, e si venga a creare un meccanismo inibitorio che vincola la generalizzazione della costruzione a nuovi elementi lessicali.

Emerge tuttavia un'eccezione interessante. Fra gli stimoli selezionati per la categoria $H_{(\text{rompere})} - L_{(\text{infrangere})}$, abbiamo testato le espressioni *rompere gli schemi* (H) e *infrangere gli schemi* (L). Il t-test compiuto sui giudizi di accettabilità raccolti per entrambe le espressioni, riporta un p-value pari a 0,2489, indice di una differenza non significativa. Il risultato è particolarmente interessante poiché l'espressione *rompere gli schemi*, costituisce un'espressione idiomatica della lingua italiana, cosa che lascia presupporre una certa fissità e rigidità semantica e sintattica che invece, nel caso specifico, non sembra verificarsi. In linea con quanto abbiamo già detto per le espressioni *ingoiare un rospo* e *inghiottire un rospo*, tale risultato va a favore di un'immagine della produttività

come fenomeno pervasivo della lingua, che riguarda costruzioni a vari livelli di schematicità e come essa non escluda anche costruzioni di natura idiomatica.

Il grafico sottostante riassume l'andamento dei giudizi di accettabilità per le varie condizioni a confronto.

Figura 36 Confronto medie dei giudizi di accettabilità espressi per *infrangere* e *rompere*.



Colmare-riempire.

Questa coppia di verbi presenta un andamento simile a quella appena analizzata.

Dal t-test emerge che non vi è differenza significativa fra i giudizi espressi per le espressioni appartenenti alle categorie H per entrambi i verbi (p-value= 0,4849): esistono una serie di nomi che ricorrono con entrambi e il livello di accettabilità degli stessi con l'uno o con l'altro non cambia.

Per quanto riguarda la categoria $H_{(colmare)} - L_{(riempire)}$, non vi è differenza significativa (p-value= 0,0808); questo significa che i soggetti sono portati a generalizzare gli usi di *colmare* a *riempire* anche se di questi ultimi il corpus presenta basse attestazioni.

Per le altre categorie i risultati sono sempre significativi: $L_{(colmare)} - H_{(riempire)}$ (p-value= 3,84007E-07), $A_{(colmare)} - Z_{(riempire)}$ (p-value= 4,75332E-07), $Z_{(colmare)} - A_{(riempire)}$ (p-value= 2,11537E-15).

L'interpretazione che possiamo fornire in questo caso è la seguente: come nel caso precedente, le differenze nella selezione argomentale riguardano, per quanto concerne *colmare*, da una parte usi idiomatici, cristallizzati nell'uso (espressioni come *colmare la misura*) dall'altra, usi di carattere tendenzialmente collocazionale (*colmare un divario*, *colmare una distanza*, *colmare una carenza*, *colmare una mancanza*) in cui fra nome e verbo si instaura un legame forte di attrazione reciproca che porta ad un radicamento dell'espressione nella mente del parlante e ad una conseguente diminuzione di

produttività. Inoltre, nel capitolo precedente avevamo visto come gli argomenti appartenenti ai primi ranghi, in un ordine di frequenza di ricorrenza con il verbo, discendente, riportassero valori molto alti e come questi ultimi, subissero poi un calo molto repentino, con gli altri nomi. Questo spiega come mai *colmare* non risulta generalizzabile rispetto agli usi di *riempire*. Gli argomenti più frequenti, presentano un grado di radicamento con il pattern, più elevato in virtù del ripetuto processo di co-attivazione fra verbo e nome. Questo porta, inevitabilmente, la costruzione ad essere associata a questi argomenti e ad una sua minore predisposizione ad essere estesa a nuovi elementi lessicali.

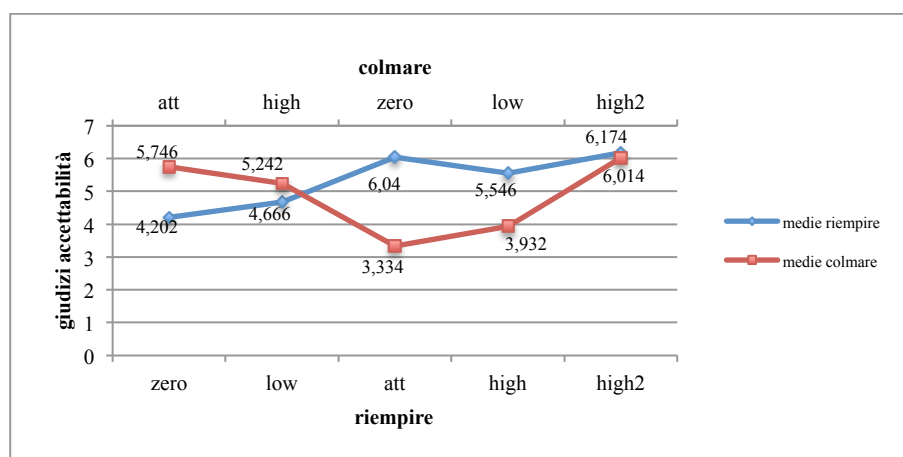
La mancata generalizzazione di *riempire* agli usi di *colmare* è più particolare e meno immediata. Abbiamo visto che nella condizione $L_{(riempire)} - H_{(colmare)}$, vi è un'effettiva generalizzazione degli usi di *colmare* a *riempire*, cosa che invece non si verifica nella condizione $Z_{(riempire)} - A_{(colmare)}$, in cui la differenza nei giudizi espressi è netta.

È probabile che in questa direzione abbia agito anche un fenomeno di *preemption* statistica: il fatto che i parlanti siano sottoposti costantemente a espressioni costituite da determinati nomi (*deficit, distanza, gap, carenza, perdita*) a fianco del verbo *colmare*, costituisca un vincolo alla generalizzazione degli stessi al verbo sinonimico; ipotesi che necessita di ulteriori indagini e accertamenti.

Possiamo vedere come si verifichi una sorta di scalarità del fenomeno che porta il verbo *riempire* ad una generalizzazione progressiva a partire dagli usi non attestati in cui il fenomeno non si verifica, a un'incipiente produttività nella condizione $L_{(riempire)} - H_{(colmare)}$, fino a una effettiva interscambiabilità dei due verbi con i nomi con cui sono attestati in larga misura nel corpus.

L'andamento dei giudizi di accettabilità rispetto alle condizioni prese in analisi è riassumibile nel grafico riportato qui sotto.

Figura 37 Confronto medie dei giudizi di accettabilità espressi per *colmare* e *riempire*.



Sollezare-alzare.

Il t-test compiuto sui giudizi di accettabilità nelle varie condizioni a confronto, per i verbi *sollevare* e *alzare*, ha portato i seguenti risultati.

Nella condizione $H_{(sollevare)} - L_{(alzare)}$, il test non riporta una differenza significativa fra i due gruppi ($p\text{-value} = 0,7092$), a dimostrazione che esiste un insieme di nomi che accettano nella stessa misura un verbo o l'altro.

Nella condizione $H_{(sollevare)} - L_{(alzare)}$, vi è una differenza nettamente significativa ($p\text{-value} = 0,00083$). I soggetti non sembrano portati a generalizzare gli argomenti più frequenti di *sollevare*, ad *alzare*, nonostante gli stessi nomi siano attestati nel corpus anche con quest'ultimo, seppure in misura minore.

È necessario fare una precisazione. Un po' come per la coppia *imparare-apprendere*, anche in questo caso, *alzare* e *sollevare*, presentano un significato comune, quello di "portare, levare in alto", entrambi però presentati in sensi secondari: *alzare*, nel senso di "aumentare", "rendere più alto" e *sollevare*, come "far sorgere, causare". Nella scelta dei nomi da utilizzare come stimoli, abbiamo escluso il secondo senso di *alzare* perché mai condiviso con *sollevare* ma non il caso contrario. Il corpus, infatti, riportava attestazioni, seppur con bassa frequenza, di espressioni come *alzare una protesta*, *alzare un vespaio*, che riflettono l'utilizzo del verbo *alzare* proprio con il significato di "far sorgere". Abbiamo per questo motivo, deciso di testare la predisposizione dei parlanti a generalizzare il verbo *alzare* anche a questi usi, con cui ha scarsa attestazione. I risultati sembrano mostrare una mancata predisposizione dei soggetti in tale direzione. Se confrontiamo i giudizi di accettabilità raccolti per le espressioni *sollevare una protesta* vs. *alzare una protesta*, otteniamo una differenza nettamente significativa (p -

value= 0,0017), mentre lo stesso confronto per le espressioni *sollevare un vespaio* e *alzare un vespaio*, riporta un t-test ai limiti della significatività (p-value= 0,0456), il che potrebbe far ipotizzare una tendenza di generalizzazione in tal senso. È probabile però, che alle due espressioni, in mancanza di contesto, i parlanti abbiano dato due letture differenti, nel primo caso un'interpretazione figurata, nel senso di “suscitare un putiferio”, nel secondo una lettura più letterale, “levare un nido di vespe”.

Sarebbe interessante compiere ulteriori indagini in questo senso, sia aumentando il numero di giudizi raccolti, sia presentando le medesime espressioni all'interno di un adeguato contesto linguistico che possa aiutare l'interpretazione delle stesse nel senso desiderato: *la questione ha sollevato un vespaio* vs. *la questione ha alzato un vespaio*.

Nella condizione $L_{(sollevare)} - H_{(alzare)}$, il t-test non riscontra una differenza significativa (p-value= 0,6030): i nomi che nel corpus erano attestati con frequenza elevata con *alzare*, ma erano poco frequenti con *sollevare*, sono accettati dai parlanti nella stessa misura con entrambi i verbi.

Nelle condizioni $A_{(sollevare)} - Z_{(alzare)}$ e $Z_{(sollevare)} - A_{(alzare)}$, il t-test riflette, nel primo caso, una differenza significativa nella distribuzione dei giudizi espressi per le due categorie messe a confronto (p-value= 1,37773E-16), nel secondo caso nessuna differenza significativa (p-value= 0,84226).

Questo è in linea con quanto detto sopra. I nomi che avevamo selezionato per la condizione $A_{(sollevare)}$, facevano capo al significato secondario di *sollevare*, nel senso di “far sorgere”, lo scopo era quello di verificare se *alzare* fosse generalizzabile in questa direzione. I risultati mostrano una risposta negativa.

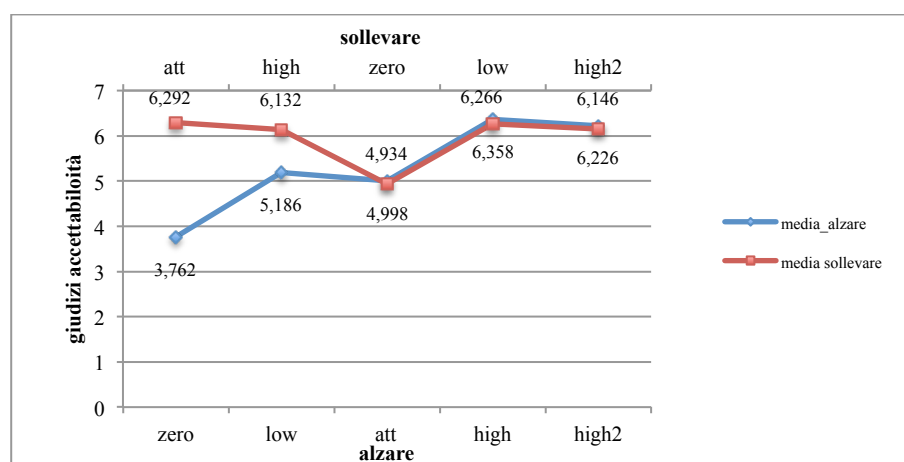
Nel caso contrario, invece, notiamo come i nomi che trovavano attestazione esclusivamente con *alzare* siano accettati nella stessa misura anche come argomenti di *sollevare*.

Un'altra precisazione è opportuna. Nel capito 3, abbiamo visto come entrambi i verbi pur essendo caratterizzati da un'elevata frequenza token, presentassero un vocabolario poco ricco e abbastanza ripetitivo e come questo si riflettesse in indici bassi di TTR (*sollevare*+N, 8,87%; *alzare*+N, 5,67%). Avevamo notato, inoltre che *alzare*+N, pur presentando un numero di attestazioni totali più elevato di *sollevare*+N, fosse caratterizzato da un numero di types e di hapax legomena più basso, motivo per cui avevamo ipotizzato una sua minore predisposizione ad essere esteso a nuovi elementi lessicali.

Questi dati sono in linea con quanto afferma Zeldes riguardo alla produttività. Egli sostiene, infatti, un meccanismo di correlazione – dissociazione fra produttività e radicamento: il fatto che una costruzione sia ben radicata non implica che sia automaticamente più produttiva, la generalizzazione di una costruzione a nuovi elementi lessicali dipende anche dal grado di radicamento dei nomi che sono istanza del suo slot argomentale. Esistono infatti costruzioni che pur essendo molto comuni nella lingua, presentano un vocabolario molto ripetitivo.

L'andamento delle medie dei giudizi di accettabilità per le varie condizioni a confronto, è riassumibile dal grafico.

Figura 38 Confronto medie dei giudizi di accettabilità espressi per *sollevare* e *alzare*.



Possiamo vedere una netta divergenza nelle condizioni $A_{(sollevare)} - Z_{(alzare)}$ e $H_{(sollevare)} - L_{(alzare)}$, condizioni nelle quali i due verbi sono stati messi a confronto nella selezione argomentale che riguarda anche la divergenza di significato che intercorre fra i due, nel senso delineato sopra.

4.3 Discussione.

Complessivamente, dall'analisi dei risultati emergono alcuni punti interessanti. Per prima cosa, scaturisce un'immagine della produttività come fenomeno scalare che riguarda le costruzioni appartenenti a vari livelli di schematicità.

Esistono costruzioni che presentano una maggiore predisposizione di altre a essere estese a nuovi elementi lessicali o a elementi lessicali rari. Il fenomeno della produttività riguarda anche quelle costruzioni che tradizionalmente sono considerate più fisse. Abbiamo visto, infatti, come alcune espressioni idiomatiche presentino una maggiore predisposizione ad essere estese ai verbi sinonimici (*ingoiare il rospo* → *inghiottire il rospo*; *rompere gli schemi* → *infrangere gli schemi*) rispetto ad altre (*calmare i bollori* → *placare i bollori*).

Dai risultati ottenuti, emerge chiaramente che nel determinare la produttività di una costruzione, intervengano numerosi fattori.

La componente semantica gioca un ruolo fondamentale nel guidare i parlanti ad accettare un'espressione piuttosto che un'altra. Lo dimostrano i risultati ottenuti presentando ai soggetti espressioni costituite dai verbi *inghiottire* e *ingoiare* seguiti da argomenti prototipicamente non ingeribili: *una strada, una casa, un uomo*, nomi appartenenti alla classe [-ingeribile].

La componente semantica tuttavia non costituisce l'unico fattore discriminante; Numerosi sono gli elementi che guidano una costruzione ad accettare nuovo materiale lessicale all'interno dei propri slot argomentali: fattori specifici della lingua e delle costruzioni stesse interagiscono nel vincolare la produttività, la quale costituisce un fenomeno arbitrario e imprevedibile, strettamente legato agli usi linguistici.

In linea con un orientamento usage-based di analisi del linguaggio abbiamo individuato nella frequenza un elemento fondamentale nello stabilire il grado di confidenza che un parlante ha con una determinata espressione, e la sua relativa predisposizione ad estenderla a nuovi utilizzi.

Dai risultati, emerge l'importanza svolta dalla frequenza type, ovvero dal numero di argomenti diversi con cui ricorre un determinato verbo. Nella fattispecie, abbiamo notato che i verbi che all'interno delle coppie di quasi-sinonimi presentavano, nel corpus, un vocabolario più ampio o un indice di TTR più elevato, sono anche quelli più predisposti ad essere generalizzati a nuovi usi.

Un ruolo fondamentale è svolto anche dalla forza del legame che si instaura fra un verbo e i suoi argomenti, attraverso l'uso che di un pattern è fatto all'interno di una lingua. Abbiamo visto che i verbi che stringono legami particolarmente robusti con i propri argomenti sono anche quelli meno soggetti al fenomeno della generalizzazione (si vedano casi come *colmare, infrangere, rompere*). In generale, gli argomenti più frequenti tendono a stringere con il verbo legami più forti, in virtù di un processo di

attivazione reciproca che si verifica ogni volta che sono utilizzati insieme; si istaura così un meccanismo inibitorio, che vincola il processo di generalizzazione del pattern. Gli argomenti più rari, al contrario istaurano legami più deboli e non esclusivi: in questo modo, gli slot argomentali che sono attestati con un numero elevato di argomenti rari (hapax legomena) o di types, hanno una maggiore predisposizione ad attrarre nuovi elementi lessicali.

Si delinea un'interazione fra il radicamento di una costruzione, scaturito dalla frequenza con cui essa è utilizzata nella lingua, e il radicamento dei suoi argomenti.

Il primo di per sé non è un fattore indicativo della produttività di un pattern: esistono, infatti, costruzioni molto comuni ma che presentano un vocabolario ripetitivo (*condurre*+N, *guidare*+N, *alzare*+N, *sollevare*+N), così come costruzioni rare che presentano un vocabolario ricco e vario (*inghiottire*+N, *ingoiare*+N).

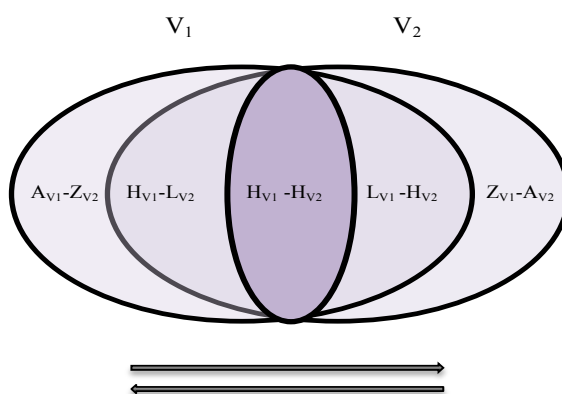
Inoltre, abbiamo visto che, nonostante alcuni verbi vengano riportati dai vocabolari come sinonimi, tale sinonimia, talvolta non è colta dai parlanti, i quali al contrario, percepiscono una differenza nell'utilizzo di un verbo, rispetto a quell'altro rispetto agli argomenti selezionati. È il caso di *condurre* e *guidare* per i quali è stata registrata una differenza nei giudizi di accettabilità espressi dai soggetti, quando i due verbi gli sono stati presentati accompagnati da nomi con i quali entrambi presentavano un'elevata attestazione nel corpus. Si è verificata anche la situazione opposta: *adempiere* e *ottemperare* pur presentando due significati diversi, sono stati percepiti dai parlanti come interscambiabili in quasi tutte le condizioni.

Un altro fattore importante che è emerso dall'analisi compiuta è che, all'interno delle coppie di sinonimi, il fenomeno della generalizzazione, quando si verifica, può avvenire in varie direzioni.

Una generalizzazione bidirezionale per cui gli usi di un verbo sono generalizzati agli usi dell'altro e viceversa. È il caso di *adempiere* e *ottemperare*. In questa coppia, sono accettati nella stessa misura²⁷, per entrambi i verbi, gli argomenti che ricorrevano, nel corpus, con bassa frequenza con un verbo ma alta con quell'altro (e viceversa), e i nomi che erano attestati solo con un verbo ma non con l'altro e viceversa.

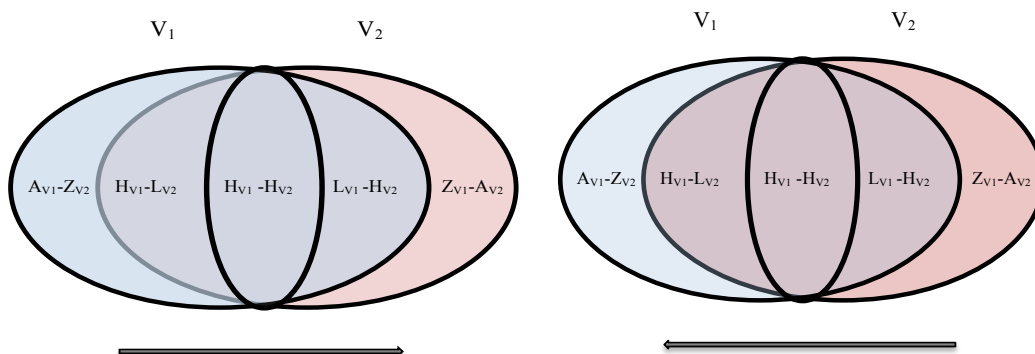
²⁷ Ad eccezione della categoria $H_{(adempiere)} - L_{(ottemperare)}$.

Figura 39 Generalizzazione bidirezionale.



Una generalizzazione unidirezionale: per cui solo gli usi di un verbo vengono generalizzati all'altro membro della coppia ma non si verifica la situazione opposta. In questo caso, uno dei due verbi presenta dunque un grado maggiore di produttività e quindi una predisposizione più elevata a essere esteso a nuovi elementi lessicali: caso di *cominciare-avviare*.

Figura 40 Generalizzazione unidirezionale.



Esistono però una serie di situazioni intermedie e intersecate: casi in cui un verbo della coppia viene esteso esclusivamente a nomi con i quali non è attestato nel corpus, ma che sono invece attestati con il verbo sinonimico; e casi in cui si assiste ad una generalizzazione progressiva degli argomenti più frequenti di uno dei due membri, anche all'altro.

4.4 Prospettive future.

Il presente studio, costituisce una sorta di indagine apripista, volta ad analizzare il fenomeno della produttività sintattica all'interno di strutture argomentali della lingua italiana. Come tale, presenta alcuni limiti e il desiderio di essere affinata per studiare il fenomeno più nel dettaglio, nelle sue varie sfaccettature.

Alcuni punti che desideriamo migliorare riguardano l'aspetto metodologico.

Il numero dei soggetti interrogati (15 partecipanti) non è molto elevato, nonostante abbiamo ritenuto che fosse sufficiente per delineare delle tendenze circa il fenomeno preso in analisi; sarebbe interessante riproporre il task estendendolo ad un maggior numero di partecipanti.

Inoltre, avendo utilizzato una metodologia crowdsourcing, per implementare il task e raccogliere i giudizi, non è stato possibile controllare l'età dei partecipanti, il luogo da cui provenivano, né il livello di istruzione, fattori questi che possono aver influenzato notevolmente le risposte ottenute. Sarebbe interessante e forse opportuno ripetere l'esperimento in ambito laboratoriale in modo da avere un maggiore controllo su questi fattori; questo ci permetterebbe anche di comparare i risultati ottenuti attraverso le due metodologie di indagine e, di conseguenza, avere un feedback circa la validità della metodologia crowdsourcing nella raccolta di dati psicolinguistici.

Un altro aspetto da affinare riguarda la scelta degli stimoli: per ogni categoria si potrebbe selezionare un maggior numero di nomi (10) per ogni condizione, e aggiungere come ulteriore categoria di indagine, la condizione $L_{v1} - L_{v2}$, per verificare la predisposizione dei due verbi a essere estesi agli argomenti rari, con particolare attenzione agli hapax legomena. Inoltre, abbiamo visto che in alcuni casi, il contesto linguistico in cui un'espressione è inserita, può costituire un fattore discriminante nell'esito delle risposte. Sarebbe dunque da prendere in considerazione l'ipotesi di ampliare le espressioni fornendo un minimo contesto linguistico che possa facilitare, ai soggetti, la comprensione delle stesse.

In un possibile follow-up del presente studio sarebbe interessante introdurre come fattore di indagine la *similarità*. Per ogni nome che ricorre nel corpus con un verbo, sarebbe interessante ricercare, attraverso metodi distribuzionali, possibili sinonimi che con quest'ultimo non sono attestati. In questo modo sarebbe possibile indagare il ruolo della *similarità semantica* nel fenomeno della generalizzazione, fattore a cui gli studi precedenti hanno attribuito vitale importanza.

Infine, come già suggerito da Zeldes (2012), sarebbe interessante misurare i tempi di reazione in un task di decisione lessicale (*lexical decision task*), nel quale ai soggetti sono presentate espressioni costituite da elementi lessicali attestati e non attestati, comuni o rari, sulla base di stime di frequenza ricavate da corpus.

In un task del genere, ci potremmo aspettare un minor tempo di reazione da parte dei soggetti nell'elaborare elementi lessicali nuovi, all'interno di slot argomentali produttivi; mentre un tempo di reazione più lungo nel caso contrario.

Come abbiamo avuto modo di vedere lungo tutto il lavoro compiuto, la produttività costituisce un ambito di indagine interessante e molto sfaccettato; numerosi sono i fattori da prendere in considerazione e l'interazione che intercorre fra essi è complessa.

Questo studio si è riproposto di analizzarne alcuni e i risultati che ne sono emersi lasciano adito all'ipotesi e alla speranza che ulteriori studi possano essere compiuti in questa direzione e possano portare a risultati interessanti.

Conclusioni.

Questo studio si è posto l'obiettivo di analizzare, attraverso un'indagine di carattere sperimentale, il fenomeno della produttività nella selezione argomentale all'interno di coppie di verbi italiani quasi-sinonimici.

Si inserisce in un panorama di studi psicolinguistici di orientamento usage-based, i quali riconoscono, nella spiegazione dei fenomeni linguistici, un'importanza primaria all'esperienza che i parlanti hanno con la lingua. Ogni segno linguistico, è memorizzato nel lessico mentale insieme a informazioni che riguardano l'uso che di quel particolare segno viene fatto nella lingua.

È a partire dagli *usage-events*, produzioni linguistiche concrete, che rappresentano non solo il prodotto del sistema linguistico, ma anche l'output, che i parlanti derivano il complesso sistema di costruzioni linguistiche.

La produttività stessa, costituisce parte della conoscenza linguistica dei parlanti. Informazioni circa il grado di estendibilità di una costruzione a nuovo materiale lessicale, sono memorizzate nel lessico mentale a fianco delle costruzioni stesse e a informazioni circa il loro uso e la loro distribuzione.

Il concetto di produttività è affiancato da quello di schematicità: con l'aumentare/il diminuire della schematicità, il significato di una costruzione diventa rispettivamente più generale o più specifico. Un significato più schematico può applicarsi ad uno spettro più ampio di situazioni, questo lascia adito ad un maggior numero di elementi possono diventare istanza dello schema stesso.

L'occorrenza di nuovi types all'interno degli slot della costruzione, contribuisce alla sua estensione e porta eventuali fenomeni di aggiustamento qualora il nuovo elemento non sia coperto dallo schema.

Un altro fattore di estrema importanza, è rappresentato dal grado di radicamento di una costruzione nella mente del parlante derivato dalla frequenza con cui una costruzione è utilizzata nella lingua. Di per sé esso non costituisce un indice particolarmente indicativo del livello di produttività di un pattern: abbiamo visto, infatti, come costruzioni molto comuni possano essere caratterizzate da un vocabolario scarso e ripetitivo.

Ciò che caratterizza il livello di generalizzazione di un pattern deriva invece dalla combinazione fra il radicamento dello stesso e quello dei suoi elementi lessicali (nel nostro caso, gli argomenti).

Ogni volta che una costruzione è utilizzata, le parti che al suo interno rimangono costanti, sono rafforzate (*entrenched*) nella rappresentazione mentale che il parlante ha della costruzione stessa, e si rendono disponibili come modelli analogici per nuove espressioni.

Abbiamo visto infatti, come gli argomenti più frequenti, lessicalizzati e in particolare le collocazioni, sviluppano, con i verbi che li selezionano, legami molto forti in virtù di un processo di co-attivazione reciproca che si instaura fra il verbo e il nome.

Ne deriva, che la costruzione, viene associata, dai parlanti, principalmente a questi argomenti e di conseguenza, un meccanismo inibitorio vincola la generalizzazione della stessa a nuovi elementi lessicali.

Al contrario, una costruzione caratterizzata da un elevato numero di types o di hapax legomena, presenterà una maggiore predisposizione ad attrarre nuovi elementi lessicali, poiché non ci sono argomenti dominanti che sono associati ad essa in modo esclusivo. In questo modo, gli argomenti meno frequenti, proprio per il fatto di essere poco radicati, vanno a rafforzare la rappresentazione astratta della costruzione.

L'inventario di argomenti, associati ad una costruzione nel lessico mentale, andrà da quelli lessicalizzati e altamente familiari, a una collezione di argomenti rari, occasionali. Ne scaturisce una visione della produttività come fenomeno scalare, che riguarda costruzioni a vario livello di schematicità, in cui la componente semantica, nonostante giochi un ruolo molto importante nel guidare i parlanti ad estendere pattern sintattici a nuovi utilizzi, non costituisce l'unico fattore discriminante.

Il presente studio, si è proposto di analizzare la produttività all'interno del fenomeno di selezione argomentale, andando a indagare i pattern di generalizzazione di coppie di verbi sinonimici.

Per farlo, abbiamo estratto pattern V+N, tramite Lexit, un framework computazionale per l'analisi distribuzionale di dati linguistici.

Dopo aver raccolto informazioni riguardo alla frequenza e alle tendenze di selezione di ciascun verbo, abbiamo creato un test psicolinguistico volto ad analizzare la predisposizione dei parlanti (misurata attraverso l'espressione di un giudizio di accettabilità) a generalizzare i membri della coppia di quasi sinonimi agli usi attestati, rispettivamente con l'altro membro e che con il primo non trovavano attestazione nel corpus, o la trovavano in misura ridotta.

Abbiamo voluto verificare, inoltre, se vi fosse una maggiore tendenza da parte di uno dei due, in questa direzione.

I risultati sembrano in linea con quanto delineato negli studi precedenti. Sicuramente, un ruolo importante è svolto dalla componente semantica: la generalizzazione della costruzione a nuovi elementi lessicali può avvenire purché vengano rispettati i vincoli semantici imposti dalla costruzione stessa.

Il fenomeno, però, abbraccia tutta una serie di fattori aggiuntivi che riguardano l'uso che della costruzione è fatto nella lingua. Abbiamo visto come, all'interno delle coppie di quasi sinonimi, il membro caratterizzato da un vocabolario più ricco, sia anche quello più predisposto ad essere esteso a nuovo materiale lessicale. Il radicamento degli argomenti, gioca un ruolo discriminante: i pattern caratterizzati da un forte legame di associazione reciproca fra il verbo e i propri argomenti, sono anche quelli che hanno presentato un minor livello di generalizzazione a nuovi usi.

Infine, abbiamo visto come effettivamente, la produttività costituisca un fenomeno scalare che riguarda costruzioni a vario livello di schematicità, da costruzioni più rigide e fisse a pattern più regolari e generali. Le costruzioni idiomatiche non sono escluse da questo fenomeno: alcune di esse sono state accettate in ugual misura sia quando sono state presentate ai soggetti con il loro verbo prototipico, sia quando quest'ultimo è stato sostituito dal quasi-sinonimo.

In conclusione, possiamo dire che questo lavoro rappresenta uno dei primi tentativi di studiare il fenomeno della produttività nella selezione argomentale all'interno di strutture V+N italiane. Come tale, sicuramente necessita di ulteriori indagini e approfondimenti che consentano di trattare il fenomeno nelle sue tante sfaccettature. La produttività come motore di evoluzione linguistica rappresenta una componente pervasiva di tutte le lingue e come tale è alimentata da meccanismi di interazione sottili e complessi. Tale complessità deve essere accettata come presupposto, se si vuole analizzare il fenomeno nella sua interezza.

Bibliografia.

- Ambridge, B., Pine, J. M., Rowland, C. F., Young, C. R., 2008, "The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgements of argument- structure overgeneralization errors", in *Cognition*, 106(1), pp. 87-129.
- Baayen, R. H., 2001, *Word Frequency Distributions*, in Text, Speech and Language Technologies, Vol. 18, Dordrecht, Kluwer Academic Publishers.
- Baayen, R. H., 2009, "Corpus linguistics in morphology: Morphological productivity", in Lüdeling, A., Kyöto, M., (a cura di), *Corpus Linguistics. An international Handbook*, Vol. 2, Berlino, Mouton de Gruyter, pp. 899-919.
- Bannard, C., Matthews, D., 2008, "Stored Word Sequences in Language Learning. The Effect of Familiarity on Children's Repetition of Four-Word Combinations", in *Psychological Science*, 19(3), pp. 241-248.
- Bannard, C., Matthews, D., 2010, "Children's Production of Unfamiliar Word Sequences is Predicted by Positional Variability and Latent Classes in a Large Sample of Child-Directed Speech", in *Cognitive Science*, 34(3), pp. 465-488.
- Barðdal, J., 2006, "Predicting the productivity of argument structure constructions", in *Annual Meeting Of The Berkeley Linguistics Society*, 32(1), pp. 467-478.
- Barðdal, J., 2008, *Productivity: evidence from case and argument structure construction in Icelandic*, Vol. 8, Amsterdam/Philadelphia, John Benjamins Publishing Company.
- Baroni, M., Bernardini, S., Comastri, F., Piccioni, L., Volpi, A., Aston, G., Mazzoleni, M., 2004, "Introducing the La Repubblica corpus: A large, annotated, TEI (XML)-compliant corpus of newspaper Italian", in *Issues*, 2, pp. 5-163.
- Bauer, L., 2001, *Morphological Productivity*, Vol. 95, New York/Cambridge, Cambridge University Press.

- Bencini, G., Goldberg, A. E., 2000, “The contribution of argument structure constructions to sentence meaning”, in *Journal of Memory and Language*, 43(4), pp. 640-651.

- Bloomfield, L., 1935, *Language*, London, George Allen & Unwin.

- Boas, H. C., 2008, “Determining the structure of lexical entries and grammatical constructions in Construction Grammar”, in *Annual Review of Cognitive Linguistics*, 6(1), pp. 113-144.

- Boyd, J. K., Goldberg, A. E., 2011, “ Learning what not to say: The role of statistical preemption and categorization in a-adjective production”, in *Language*, 87(1), pp. 55-83.

- Boyd J. K., Goldberg, A. E., 2012, “Young children fail to fully generalize a novel argument structure construction when exposed to the same input as older learners”, in *Journal of child language*, 39(3), pp. 457-481.

- Bybee, J. L., 2006, “From usage to grammar: The mind’s response to repetition”, in *Language*, 82(4), pp. 711-733.

- Bybee, J. L., 2010, *Language, Usage And Cognition*, New York, Cambridge University Press.

- Bybee, J. L., Eddington D., 2006, “ A Usage-Based approach to Spanish verbs of becoming”, in *Language* 82(2), pp. 323-355.

- Bybee, J. L., Thompson, S., 1997, “Three frequency effects in syntax”, in *Annual Meeting of the Berkeley Linguistics Society*, 23(1), pp. 378-388.

- Cameron-Faulkner, T., Lieven, E., Tomasello M., 2003, “A Construction Based Analysis Of Child Directed Speech”, in *Cognitive Science*, 27(6), pp. 843-873.

- Casenhiser, D., Goldberg, A. E., 2005, "Fast mapping between a phrasal form and meaning", in *Developmental Science* 8(6), pp. 500-508.

- Chomsky, N., 1957, *Syntactic Structures*, The Hague/Paris, Mouton.

- Chomsky, N., 1965, *Aspects of the Theory of Syntax*, Cambridge, MA, MIT Press.

- Clausner, T. C., Croft, W., 1997, "Productivity and schematicity in metaphors", in *Cognitive Science*, 21(3), pp. 247-282.

- Croft, W., Cruse, A. D., 2004, in *Cognitive Linguistics*, Cambridge, Cambridge University Press.

- Di Marco, C., Hirst, G., Stede, M., 1993, "The semantic and stylistic differentiation of synonyms and near-synonyms", in *AAAI Spring Symposium on Building Lexicons for Machine Translation*, pp. 114-121.

- Diessel, H., 2007, "Frequency effects in language acquisition, language use, and language diachronic change", in *New ideas in psychology*, 25(2), pp. 108-127.

- Ellis, N. C., 2002, "Frequency effects in language processing", in *Studies in second language acquisition*, 24(2), pp. 143-188.

- Fillmore, C., Kay, P., O'Connor, M. C., 1988, "Regularity and Idiomaticity in Grammatical Constructions: The case of Let Alone", in *Language*, 64(3), pp. 501-538.

- Firth, J. R., 1957, *Papers in Linguistics*, London, Oxford University Press.

- Goldberg, A. E., Bencini, G. M., 2005, "Support from language processing for a constructional approach to grammar", in *Language in use: Cognitive and discourse perspectives on language and language learning*, pp. 3-18.

- Goldberg, A. E., 1995, *Constructions: A construction grammar approach to*

argument structure, Chicago/London, University Of Chicago Press.

- Goldberg, A. E., 2006, *Constructions at work: The nature of generalization in language*, New York, Oxford University Press on Demand.
- Goldberg, A. E., 2008, "The nature of generalization in language", in *Cognitive Linguistics*, 20(1), pp. 93-127.
- Goldberg, A. E., 2011, "Corpus evidence of the viability of statistical preemption", in *Cognitive Linguistics*, 22(1), pp. 131-153.
- Goldberg, A. E., 2013, "Argument structure constructions versus lexical rules or derivational verb templates", in *Mind & Language*, 28(4), pp. 435-465.
- Goldberg, A., Casenhiser, D. M., Sethuraman, N., 2004, "Learning argument structure generalizations", in *Cognitive linguistics* 15(3), pp. 289-316.
- Healy, A., Miller, G., 1970, "The Verb as the Main Determinant of Sentence Meaning", in *Psychonomic Science*, 20, pp. 372.
- Hockett, C. F., 1958, *A Course in Modern Linguistics*, New York, Macmillan.
- Hoffmann, T., Trousdale, G., 2013, (a cura di), *The Oxford Handbook Of Construction Grammar*, New York, Oxford University Press.
- Inkpen, D., Hirst, G., 2006, "Building and using a lexical knowledge base of near-synonym differences", in *Computational linguistics*, 32(2), pp. 223-262.
- Israel, M., 1996, "The Way Constructions Grow", in *Conceptual Structure, Discourse and Language*, Goldberg, A. E., (a cura di), Standford, CSLI Publications, pp. 217-230.
- Kidd, E., Lieven, E., Tomasello, M., 2006, "Examining The Role Of Lexical Frequency In The Acquisition And Processing Of Sentential Complements", in *Cognitive Development*, 21(2), pp. 93-107.

- Langacker, R., W., 1999, *Grammar and Conceptualization*, Berlin/New York, Mouton de Gruyter.
- Lenci, A., 2008, "Distributional Semantics in Linguistic and Cognitive Research", in *Italian journal of Linguistics*, 20(1), pp.1-31.
- Lenci, A., Lapesa, G., Bonansinga G., 2012, "LexIt: A Computational Resource on Italian Argument Structure", in *LREC*, pp. 3712-3718.
- Levin B., 1993, *English Verb Classes and Alternations: a Preliminary Investigation*, Chicago/Londra, University of Chicago Press.
- Levin, B., Rappaport Hovav, M., 2005, *Argument Realization*, New York, Cambridge University Press.
- Lieven, E., Behrens, H., Speares, J., Tomasello, M., 2003, "Early syntactic creativity: A usage-based approach", in *Journal of child language*, 30(2), pp. 333-370.
- Miller, G. A., Charles, W. G., 1991, "Contextual correlates of semantic similarity", in *Language and Cognitive Processes*, 6(1), pp. 1-28.
- Müller, S., 2006, "Phrasal or lexical constructions?", in *Language*, 82(4), pp. 850-883.
- Müller, S., Wechsler, S., 2014, "Lexical approaches to argument structure", in *Theoretical Linguistics*, 40(1-2), pp. 1-76.
- Munro, R., Bethard, S., Kuperman, V., Lai, V. T., Melnick, R., Potts, C., Schnoebelen T., Tily, H., 2010, "Crowdsourcing and language studies: the new generation of linguistic data", in *Proceedings of the NAACL HLT 2010 workshop on creating speech and language data with Amazon's Mechanical Turk*, Association for Computational Linguistics, pp. 122-130.
- Murphy, M. L., 2003, *Semantic Relations and the Lexicon*, Cambridge, Cambridge University Press.

- Murphy, M. L., 2013, “What we talk about when we talk about synonyms (and what it can tell us about thesauruses)”, in *Internal Journal of Lexicography*, 26(3), pp. 279-304.
- Olofsson, J., 2014, “Argument structure constructions and syntactic productivity. The case of Swedish motion constructions”, in *Constructions*, 1(7), pp. 1-17.
- Perek, F., 2015, *Argument structure in Usage-Based Construction Grammar*, Amsterdam/Philadelphia, John Benjamins Publishing Company.
- Perek, F., 2016 (to appear), “Using distributional semantics to study syntactic productivity in diachrony: A case study”, in *Linguistics*, 54(1).
- Perek, F., Goldberg, A. E., 2015, “Generalizing beyond the input: The functions of the constructions matter”, in *Journal Of Memory And Language*, Vol. 84, pp. 108-127.
- Pinker, S., 1989, *Learnability and cognition: The acquisition of argument structure*, Cambridge, MA, MIT Press.
- Pinker, S., 2015, *Words and Rules: The Ingredients of Language*, New York, Basic Books.
- Rappaport Hovav, M., Levin, B., 1998, "Building verb meanings", in Butt, M., Geuder, W., (a cura di), *The Projection of Arguments*, Stanford, CA, CSLI Publications, pp. 97-134.
- Sahlgren, M., 2006, "The Word-Space Model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces.", Phd dissertation, Stockholm University, Faculty of Humanities, Department of Linguistics.
- Suttle, L., Goldberg, A. E., 2011, “The partial productivity of constructions as induction”, in *Linguistics*, 49(6), pp. 1237–1269.

- Taylor, J. R., 2003, "Near synonyms as co-extensive categories: 'high' and 'tall' revisited", in *Language Sciences*, 25(3), pp. 263-284.
- Tesnière, L., 1959, *Éléments de syntaxe structurale*, Parigi, Librairie Klincksieck.
- Theakston, A. L., 2004, "The role of entrenchment in children's and adults' performance on grammaticality judgment tasks", in *Cognitive Development*, 19(1), pp. 15-34.
- Theakston, A. L., Ibbotson, P., Fredenthal, D., Lieven, E. V., Tomasello, M., 2015, "Productivity Of Noun Slots In Verb Frames", in *Cognitive Science*, 39(6) pp.1369-1395.
- Tomasello, M., 2006, "Acquiring Linguistic Constructions", in *Handbook of Child Psychology*, Siegler, R., Kuhn, D., (a cura di), Vol. 2, pp. 255-298.
- Turney, P. D., Pantel, P., 2010, "From frequency to meaning: vector space models of semantics", in *Journal Of Artificial Intelligence Research*, 37(1), pp. 141-188.
- Zeldes, A., 2009, "Quantifying constructional productivity with unseen slot members", in *Proceedings of the Workshop on Computational Approaches to Linguistic Creativity*, Association for Computational Linguistics, pp. 47-54.
- Zeldes, A., 2012, *Productivity in argument selection: From morphology to syntax*, Vol. 260, Berlin/Boston, De Gruyter Mouton.
- Zeldes, A., 2013, "Productive argument selection: Is lexical semantic enough?", in *Corpus Linguistics And Linguistic Theory*, 9(2), pp. 263-291.
- Zeschel, A., 2008, "Lexical chunking effects in syntactic processing", in *Cognitive Linguistics*, 19(3), pp. 427-446.
- Zeschel, A., 2009, "What's (in) a construction? Complete inheritance vs. full-entry models", in *New Directions In Cognitive Linguistics*, Vyvyan, E., Stéphanie, P., (a

cura di), Amsterdam/Philadelphia, John Benjamins, pp. 185-200.

- Zeschel, A., 2010, “Exemplars and analogy: Semantic extension in constructional networks”, in *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, pp. 201-219.
- Zeschel, A., 2012, *Incipient productivity: A construction-based approach to linguistic creativity*, Vol. 49, Berlin/Boston, De Gruyter Mouton.

Sitografia.

- www.lexit.fileli.unipi.it
- <https://www.sketchengine.co.uk/>
- www.crowdfunder.com
- www.garzantilinguistica.it
- www.treccani.it
- dizionari.corriere.it
- www.accademiadellacrusca.it

Appendice.

Lista completa dei verbi estratti da Lexit.

Verbo	Frequenza token	Frequenza type	Numero hapax	TTR	TTR %
Impaurire	116	86	68	0,722689076	72,27%
Mormorare	159	58	44	0,364779874	36,48%
Aborrire	185	153	132	0,827027027	82,70%
Deprecare	225	179	152	0,795555556	79,56%
Estirpare	299	145	102	0,484949833	48,49%
Scordare	382	295	254	0,772251309	77,23%
Sussurrare	394	179	125	0,454314721	45,43%
Disconoscere	411	212	158	0,515815085	51,58%
Ripudiare	426	241	168	0,5657277	56,57%
Ottemperare a	441	108	63	0,244897959	24,49%
Sradicare	489	235	164	0,480572597	48,06%
Paventare	693	351	263	0,506493506	50,65%
Assolvere a	695	71	40	0,102158273	10,22%
Adempiere a	721	90	42	0,12482663	12,48%
Inghiottire	732	426	326	0,581967213	58,20%
Deplorare	752	414	301	0,550531915	55,05%
Lodare	871	470	320	0,539609644	53,96%
Scappare da	914	401	268	0,438730853	43,87%
Urlare	964	265	171	0,274896266	27,49%
Detestare	991	634	465	0,63975782	63,98%
Ingoiare	1009	487	362	0,482656095	48,27%
Ultimare	1056	365	208	0,345643939	34,56%
Incorrere in	1141	308	192	0,26993865	26,99%
Incappare in	1185	495	328	0,417721519	41,77%
Elogiare	1187	543	358	0,457455771	45,75%
Spaventare	1303	456	264	0,349961627	35,00%
Rinnegare	1306	468	295	0,358346095	35,83%
Calmare	1331	307	181	0,230653644	23,075
Odiare	1497	770	538	0,496325985	49,63%
Fuggire da	1608	592	380	0,367537313	36,75%
Placare	1614	404	243	0,250309789	25,03%
Proibire	1813	652	401	0,359624931	35,96%

Apprendere	1898	374	240	0,197049526	19,70%
Gridare	2104	362	201	0,172053232	17,21%
Imparare	3130	597	357	0,190734824	19,27%
Terminare	3482	763	418	0,219126939	21,91%
Vietare	5224	1085	592	0,207695253	20,77%
Criticare	7160	1571	831	0,219413408	21,94%
Finire	10502	1813	953	0,172633784	17,26%
Dimenticare	10937	2998	1618	0,274115388	27,41%
Temere	11265	2364	1139	0,209853529	20,99%
Comprare	13298	1925	938	0,14483381	14,48%
Cominciare	14299	1569	685	0,109727953	10,97%
Iniziare	14416	1503	684	0,104259156	10,43%
Acquistare	17044	1996	952	0,117167332	11,72%
Avviare	19607	1440	656	0,073443158	7,34%
Guidare	14632	1367	605	0,093425369	9,34%
Condurre	11816	902	475	0,07633717	7,63%
Rompere	10487	912	441	0,086964814	8,69%
Infrangere	1881	370	226	0,196703881	19,67%
Colmare	2440	262	172	0,107377049	10,73%
Riempire	6118	1006	545	0,164432821	16,44%
Sollevare	9239	820	443	0,088754194	8,87%
Alzare	12733	723	357	0,056781591	5,67%

Stimoli usati per l'esperimento.

n_espressione	espressioni	frequenza
1	inghiottire il rospo	H
2	inghiottire un boccone	H
3	inghiottire la pillola	H
4	inghiottire miliardi	H
5	inghiottire le lacrime	H
6	inghiottire una strada	H
7	inghiottire una casa	H
8	inghiottire le risorse	H
9	inghiottire una persona	H
10	inghiottire un uomo	H
11	inghiottire una sconfitta	L
12	inghiottire un veleno	L
13	inghiottire una pasticca	L
14	inghiottire un'umiliazione	L
15	inghiottire amaro	L
16	inghiottire un anello	ZERO
17	inghiottire una delusione	ZERO
18	inghiottire un documento	ZERO
19	inghiottire un'emozione	ZERO
20	inghiottire un'idea	ZERO
21	inghiottire un'auto	A
22	inghiottire una terra	A
23	inghiottire una città	A
24	inghiottire un'immagine	A
25	inghiottire un'azienda	A
26	ingoiare il rospo	H
27	ingoiare un boccone	H
28	ingoiare la pillola	H
29	ingoiare miliardi	H
30	ingoiare le lacrime	H
31	ingoiare una strada	L
32	ingoiare una casa	L
33	ingoiare le risorse	L
34	ingoiare una persona	L
35	ingoiare un uomo	L
36	ingoiare una sconfitta	H
37	ingoiare un veleno	H
38	ingoiare una pasticca	H
39	ingoiare un'umiliazione	H
40	ingoiare amaro	H
41	ingoiare un anello	A
42	ingoiare una delusione	A
43	ingoiare un documento	A

44	ingoiare un'emozione	A
45	ingoiare un'idea	A
46	ingoiare un'auto	ZERO
47	ingoiare una terra	ZERO
48	ingoiare una città	ZERO
49	ingoiare un'immagine	ZERO
50	ingoiare un'azienda	ZERO
51	incappare in un incidente	H
52	incappare in un errore	H
53	incappare in un rigore	H
54	incappare in un reato	H
55	incappare in un infortunio	H
56	incappare in una multa	L
57	incappare in un equivoco	L
58	incappare in una violazione	L
59	incappare in un provvedimento	L
60	incappare nella scomunica	L
61	incappare in una legge	H
62	incappare in una sconfitta	H
63	incappare in un blocco	H
64	incappare in un controllo	H
65	incappare in una disavventura	H
66	incappare in una rete	A
67	incappare in un'inchiesta	A
68	incappare in una vicenda	A
69	incappare in una pattuglia	A
70	incappare in uno scandalo	A
71	incappare in una perdita	ZERO
72	incappare in un'omissione	ZERO
73	incappare in responsabilità giuridiche	ZERO
74	incappare in una penalizzazione	ZERO
75	incappare in un delitto	ZERO
76	incorrere in un incidente	H
77	incorrere in un errore	H
78	incorrere in un rigore	H
79	incorrere in un reato	H
80	incorrere in un infortunio	H
81	incorrere in una multa	H
82	incorrere in un equivoco	H
83	incorrere in una violazione	H
84	incorrere in un provvedimento	H
85	incorrere nella scomunica	H
86	incorrere in una legge	L
87	incorrere in una sconfitta	L
88	incorrere in un blocco	L
89	incorrere in un controllo	L
90	incorrere in una disavventura	L

91	incorrere in una rete	ZERO
92	incorrere in un'inchiesta	ZERO
93	incorrere in una vicenda	ZERO
94	incorrere in una pattuglia	ZERO
95	incorrere in uno scandalo	ZERO
96	incorrere in una perdita	A
97	incorrere in un'omissione	A
98	incorrere in responsabilità giuridiche	A
99	incorrere in una penalizzazione	A
100	incorrere in un delitto	A
101	imparare una lezione	H
102	imparare un segreto	H
103	imparare una lingua	H
104	imparare un'arte	H
105	imparare una tecnica	H
106	imparare una parola	H
107	imparare il rispetto	H
108	imparare l'italiano	H
109	imparare un valore	H
110	imparare un lavoro	H
111	imparare la verità	L
112	imparare un particolare	L
113	imparare una notizia	L
114	imparare un dettaglio	L
115	imparare un'informazione	L
116	imparare la lotta	A
117	imparare la democrazia	A
118	imparare una battuta	A
119	imparare il confronto	A
120	imparare la strada	A
121	imparare un principio	ZERO
122	imparare la sintassi	ZERO
123	imparare una competenza	ZERO
124	imparare un'idea	ZERO
125	imparare una novità	ZERO
126	apprendere una lezione	H
127	apprendere un segreto	H
128	apprendere una lingua	H
129	apprendere un'arte	H
130	apprendere una tecnica	H
131	apprendere una parola	L
132	apprendere il rispetto	L
133	apprendere l'italiano	L
134	apprendere un valore	L
135	apprendere un lavoro	L
136	apprendere la verità	H
137	apprendere un particolare	H

138	apprendere una notizia	H
139	apprendere un dettaglio	H
140	apprendere un'informazione	H
141	apprendere la lotta	ZERO
142	apprendere la democrazia	ZERO
143	apprendere una battuta	ZERO
144	apprendere il confronto	ZERO
145	apprendere la strada	ZERO
146	apprendere un principio	A
147	apprendere la sintassi	A
148	apprendere una competenza	A
149	apprendere un'idea	A
150	apprendere una novità	A
151	cominciare un lavoro	H
152	cominciare un processo	H
153	cominciare una campagna	H
154	cominciare un'attività	H
155	cominciare una discussione	H
156	cominciare una procedura	L
157	cominciare una riflessione	L
158	cominciare un'iniziativa	L
159	cominciare un progetto	L
160	cominciare un procedimento	L
161	cominciare una giornata	H
162	cominciare un viaggio	H
163	cominciare uno sciopero	H
164	cominciare una partita	H
165	cominciare una storia	H
166	cominciare un ripensamento	ZERO
167	cominciare una razionalizzazione	ZERO
168	cominciare una rogatoria	ZERO
169	cominciare un riassetto	ZERO
170	cominciare una creazione	ZERO
171	cominciare un allenamento	A
172	cominciare un romanzo	A
173	cominciare un'arringa	A
174	cominciare una lite	A
175	cominciare una rassegna	A
176	avviare un lavoro	H
177	avviare un processo	H
178	avviare una campagna	H
179	avviare un'attività	H
180	avviare una discussione	H
181	avviare una procedura	H
182	avviare una riflessione	H
183	avviare un'iniziativa	H
184	avviare un progetto	H

185	avviare un procedimento	H
186	avviare una giornata	L
187	avviare un viaggio	L
188	avviare uno sciopero	L
189	avviare una partita	L
190	avviare una storia	L
191	avviare un ripensamento	A
192	avviare una razzionalizzazione	A
193	avviare una rogatoria	A
194	avviare un riassetto	A
195	avviare una creazione	A
196	avviare un allenamento	ZERO
197	avviare un romanzo	ZERO
198	avviare un'arringa	ZERO
199	avviare una lite	ZERO
200	avviare una rassegna	ZERO
201	calmare l'animo	H
202	calmare le acque	H
203	calmare una tensione	H
204	calmare l'inquietudine	H
205	calmare la rabbia	H
206	calmare una polemica	L
207	calmare la fame	L
208	calmare un'irritazione	L
209	calmare una protesta	L
210	calmare la sete	L
211	calmare un mercato	H
212	calmare i nervi	H
213	calmare una situazione	H
214	calmare i bollori	H
215	calmare la gente	H
216	calmare un gioco	A
217	calmare un giocatore	A
218	calmare il mare	A
219	calmare i pensieri	A
220	calmare il cuore	A
221	calmare le critiche	ZERO
222	calmare il clima	ZERO
223	calmare i sospetti	ZERO
224	calmare il risentimento	ZERO
225	calmare i dubbi	ZERO
226	placare l'animo	H
227	placare le acque	H
228	placare una tensione	H
229	placare l'inquietudine	H
230	placare la rabbia	H
231	placare una polemica	H

232	placare la fame	H
233	placare un'irritazione	H
234	placare una protesta	H
235	placare la sete	H
236	placare un mercato	L
237	placare i nervi	L
238	placare una situazione	L
239	placare i bollori	L
240	placare la gente	L
241	placare un gioco	ZERO
242	placare un giocatore	ZERO
243	placare il mare	ZERO
244	placare i pensieri	ZERO
245	placare il cuore	ZERO
246	placare le critiche	A
247	placare il clima	A
248	placare i sospetti	A
249	placare il risentimento	A
250	placare i dubbi	A
251	adempiere a un obbligo	H
252	adempiere a un impegno	H
253	adempiere a un dovere	H
254	adempiere a una condizione	H
255	adempiere a una richiesta	H
256	adempiere a un compito	H
257	adempiere a un ruolo	H
258	adempiere a un atto	H
259	adempiere a una formalità	H
260	adempiere a una promessa	H
261	adempiere a un'ordinanza	L
262	adempiere a una regola	L
263	adempiere a una risoluzione	L
264	adempiere a una legge	L
265	adempiere a una sentenza	L
266	adempiere a una funzione	A
267	adempiere a una missione	A
268	adempiere a un ufficio	A
269	adempiere a un'obbligazione	A
270	adempiere a una mansione	A
271	adempiere a un invito	ZERO
272	adempiere a un accordo	ZERO
273	adempiere a un decreto	ZERO
274	adempiere a un divieto	ZERO
275	adempiere a una delibera	ZERO
276	ottemperare a un obbligo	H
277	ottemperare a un impegno	H
278	ottemperare a un dovere	H

279	ottemperare a una condizione	H
280	ottemperare a una richiesta	H
281	ottemperare a un compito	L
282	ottemperare a un ruolo	L
283	ottemperare a un atto	L
284	ottemperare a una formalità	L
285	ottemperare a una promessa	L
286	ottemperare a un'ordinanza	H
287	ottemperare a una regola	H
288	ottemperare a una risoluzione	H
289	ottemperare a una legge	H
290	ottemperare a una sentenza	H
291	ottemperare a una funzione	ZERO
292	ottemperare a una missione	ZERO
293	ottemperare a un ufficio	ZERO
294	ottemperare a un'obbligazione	ZERO
295	ottemperare a una mansione	ZERO
296	ottemperare a un invito	A
297	ottemperare a un accordo	A
298	ottemperare a un decreto	A
299	ottemperare a un divieto	A
300	ottemperare a una delibera	A
301	rompere un equilibrio	H
302	rompere un muro	H
303	rompere un tabù	H
304	rompere le regole	H
305	rompere una barriera	H
306	rompere gli schemi	H
307	rompere un fronte	H
308	rompere un'alleanza	H
309	rompere un accerchiamento	H
310	rompere un incantesimo	H
311	rompere un sogno	L
312	rompere un divieto	L
313	rompere un mito	L
314	rompere un limite	L
315	rompere un record	L
316	rompere un contratto	A
317	rompere l'assedio	A
318	rompere una relazione	A
319	rompere un isolamento	A
320	rompere una trattativa	A
321	rompere una legge	ZERO
322	romperere una norma	ZERO
323	rompere un ordine	ZERO
324	rompere una speranza	ZERO
325	rompere un regolamento	ZERO

326	infrangere un equilibrio	H
327	infrangere un muro	H
328	infrangere un tabù	H
329	infrangere le regole	H
330	infrangere una barriera	H
331	infrangere gli schemi	L
332	infrangere un fronte	L
333	infrangere un'alleanza	L
334	infrangere un accerchiamento	L
335	infrangere un incantesimo	L
336	infrangere un sogno	H
337	infrangere un divieto	H
338	infrangere un mito	H
339	infrangere un limite	H
340	infrangere un record	H
341	infrangere un contratto	ZERO
342	infrangere l'assedio	ZERO
343	infrangere una relazione	ZERO
344	infrangere un isolamento	ZERO
345	infrangere una trattativa	ZERO
346	infrangere una legge	A
347	infrangerere una norma	A
348	infrangere un ordine	A
349	infrangere una speranza	A
350	infrangere un regolamento	A
351	guidare un paese	H
352	guidare un'operazione	H
353	guidare un processo	H
354	guidare una politica	H
355	guidare una campagna	H
356	guidare la nazionale	H
357	guidare un'auto	H
358	guidare una rivolta	H
359	guidare una coalizione	H
360	guidare un movimento	H
361	guidare una guerra	L
362	guidare un affare	L
363	guidare un esperimento	L
364	guidare un negoziato	L
365	guidare un sondaggio	L
366	guidare un governo	A
367	guidare una delegazione	A
368	guidare una società	A
369	guidare una protesta	A
370	guidare un fronte	A
371	guidare una vita	ZERO
372	guidare un programma	ZERO

373	guidare uno studio	ZERO
374	guidare un'istruttoria	ZERO
375	guidare un interrogatorio	ZERO
376	condurre un paese	H
377	condurre un'operazione	H
378	condurre un processo	H
379	condurre una politica	H
380	condurre una campagna	H
381	condurre la nazionale	L
382	condurre un'auto	L
383	condurre una rivolta	L
384	condurre una coalizione	L
385	condurre un movimento	L
386	condurre una guerra	H
387	condurre un affare	H
388	condurre un esperimento	H
389	condurre un negoziato	H
390	condurre un sondaggio	H
391	condurre un governo	ZERO
392	condurre una delegazione	ZERO
393	condurre una società	ZERO
394	condurre una protesta	ZERO
395	condurre un fronte	ZERO
396	condurre una vita	A
397	condurre un programma	A
398	condurre uno studio	A
399	condurre un'istruttoria	A
400	condurre un interrogatorio	A
401	colmare un vuoto	H
402	colmare una lacuna	H
403	colmare un buco	H
404	colmare uno spazio	H
405	colmare un silenzio	H
406	colmare un divario	H
407	colmare una voragine	H
408	colmare un distacco	H
409	colmare una mancanza	H
410	colmare uno scarto	H
411	colmare una pagina	L
412	colmare uno stadio	L
413	colmare una piazza	L
414	colmare un cuore	L
415	colmare l'aria	L
416	colmare un deficit	A
417	colmare una distanza	A
418	colmare una differenza	A
419	colmare una carenza	A

420	colmare una perdita	A
421	colmare la bocca	ZERO
422	colmare una sala	ZERO
423	colmare una strada	ZERO
424	colmare un modulo	ZERO
425	colmare una cassa	ZERO
426	riempire un vuoto	H
427	riempire una lacuna	H
428	riempire un buco	H
429	riempire uno spazio	H
430	riempire un silenzio	H
431	riempire un divario	L
432	riempire una voragine	L
433	riempire un distacco	L
434	riempire una mancanza	L
435	riempire uno scarto	L
436	riempire una pagina	H
437	riempire uno stadio	H
438	riempire una piazza	H
439	riempire un cuore	H
440	riempire l'aria	H
441	riempire un deficit	ZERO
442	riempire una distanza	ZERO
443	riempire una differenza	ZERO
444	riempire una carenza	ZERO
445	riempire una perdita	ZERO
446	riempire la bocca	A
447	riempire una sala	A
448	riempire una strada	A
449	riempire un modulo	A
450	riempire una cassa	A
451	sollevare un polverone	H
452	sollevare un velo	H
453	sollevare un coperchio	H
454	sollevare la testa	H
455	sollevare la polvere	H
456	sollevare una protesta	H
457	sollevare un vespaio	H
458	sollevare il morale	H
459	sollevare un'ondata	H
460	sollevare un peso	H
461	sollevare il telefono	L
462	sollevare un calice	L
463	sollevare una bandiera	L
464	sollevare le spalle	L
465	sollevare il pugno	L
466	sollevare un problema	A

467	sollevare un dubbio	A
468	sollevare una polemica	A
469	sollevare un'obiezione	A
470	sollevare un interrogativo	A
471	sollevare una palla	ZERO
472	sollevare un elicottero	ZERO
473	sollevare una vela	ZERO
474	sollevare una pretesa	ZERO
475	sollevare il cartellino	ZERO
476	alzare un polverone	H
477	alzare un velo	H
478	alzare un coperchio	H
479	alzare la testa	H
480	alzare la polvere	H
481	alzare una protesta	L
482	alzare un vespaio	L
483	alzare il morale	L
484	alzare un'ondata	L
485	alzare un peso	L
486	alzare il telefono	H
487	alzare un calice	H
488	alzare una bandiera	H
489	alzare le spalle	H
490	alzare il pugno	H
491	alzare un problema	ZERO
492	alzare un dubbio	ZERO
493	alzare una polemica	ZERO
494	alzare un'obiezione	ZERO
495	alzare un interrogativo	ZERO
496	alzare una palla	A
497	alzare un elicottero	A
498	alzare una vela	A
499	alzare una pretesa	A
500	alzare il cartellino	A

