



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova

Dipartimento di Studi Linguistici e Letterari

Corso di Laurea Magistrale in Linguistica

Classe LM-39

Tesi di Laurea

La distribuzione del singolare e del plurale nei nomi dell'italiano: uno studio di rating e di corpus

Relatore

Dr. Chiara Zanini

Correlatore

dott.ssa Francesca Franzon

Laureando

Jason Cole

n° matr. 1082461 / LMLIN

Anno Accademico 2015 / 2016

INDICE

INTRODUZIONE	5
CAPITOLO I: Principali modelli del recupero lessicale in psicolinguistica	7
1.1 Modelli in comprensione.....	7
1.1.2 Modello Parallel Dual-Route.....	7
1.1.3 Baayen, Dijkstra, e Schreuder, (1997).....	9
1.1.4 Baayen, Burani, e Schreuder, (1996).....	11
1.2 Modelli in produzione.....	12
1.2.1 Levelt, Roelofs, e Mayer, (1999).....	12
1.2.2 Baayen, Levelt, Schreuder, e Ernestus, (2007)	16
1.2.3 Biedermann, Beyersmann, Mason, e Nickels, (2013)	18
1.2.4 Luzzati, Mondini, e Semenza, (2011).....	20
1.2.5 Biedermann, Lorenz, Beyersmann, e Nickels, (2012).....	21
1.2.6 Nickels, Biedermann, Fieder, e Schiller, (2015).....	22
1.2.7 Baayen, Milin, Durdevic, Hendrix, e Marelli (2011).....	25
CAPITOLO II: Uno studio quantitativo: analisi sul corpus e studio di <i>rating</i>.....	27
2.1 Selezione dei nomi nel corpus ItWaC	27
2.2 Studio di <i>rating</i>.....	29
2.3 Risultati.....	30
2.3.1 Distribuzione delle forme singolari e plurali.....	30

2.3.2 Correlazioni.....	32
2.3.3 Misure di dominanza.....	35
CAP III: Discussioni e conclusioni.....	37
Bibliografia.....	41
Appendice A.....	45

Introduzione

Questa tesi nasce con la volontà di approfondire il tema della "*plural dominance*", ossia il fenomeno per cui un nome occorre più frequentemente al plurale che al singolare. Il fenomeno è di interesse perché, in accordo con la scala di gerarchia del numero (Corbett, 2000), le forme singolari sono generalmente meno marcate e occorrono più frequentemente delle corrispondenti forme plurali. Nei casi di *plural dominance*, invece, sembrerebbe, come ipotizza Tiersma (1982), che la condizione meno marcata sia rappresentata dalla forma plurale del termine: ovvero in questi casi il plurale diverrebbe la forma più saliente del nome.

Da una prospettiva tipologica è interessante notare che quando la forma singolare di un nome è più marcata della corrispondente forma plurale alcune lingue prevedono l'aggiunta di suffissi per il singolare e non per il plurale (è il caso dei nomi singolativi). Ad esempio Bender (1983) riporta il caso del majanag, una lingua nilo-sahariana in cui per indicare il collettivo (ad es. *ηεεti* 'pidocchi') non viene utilizzata morfologia plurale, mentre nella forma singolare (*ηεεti-n* 'pidocchio') è presente suffissazione. Nomi di questo tipo sono degli esempi di plurali lessicali (vedi Acquaviva, 2008) e rappresentano un caso interessante in quanto mostrano come anche forme marginali rispetto alla lingua possano essere integrate in un sistema grammaticale. Data l'affinità tra i nomi a dominanza plurale e i nomi singolativi in questo lavoro tenteremo di chiarire se anche i nomi di tipo *plural dominant* non possano essere considerati come dei particolari tipi di plurali lessicali.

Il fenomeno della *plural dominance* in linguistica formale non ha ricevuto molta attenzione, mentre è stato maggiormente discusso in psicolinguistica. Scopo di questa tesi è quindi anche quello di collegare le nozioni provenienti dalla parte formale, astratta del linguaggio, a quelle di carattere più concreto, legate ai processi cognitivi che sottendono all'elaborazione linguistica. L'argomento della dominanza plurale offre una buona opportunità da questo punto di vista poiché in esso restano ancora aperte domande importanti quali: i nomi *plural dominant* designano categorie semantiche precise? O il verificarsi di questo fenomeno è una questione di alta probabilità legata ai contesti comunicativi con cui un dato nome occorre al plurale? Questo fenomeno, così come descritto in letteratura, corrisponde ad una competenza nella grammatica del parlante?

Per rispondere a queste domande, in questa tesi si è proceduto alla creazione di una

lista dei nomi a maggiore frequenza plurale nell'italiano, estrapolata dal corpus ItWaC. Successivamente, si è compiuto uno studio di *rating* utilizzando la medesima lista di nomi, con l'obiettivo di indagare il fenomeno della distribuzione dei singolari e dei plurali nella lingua italiana relativamente alla frequenza soggettiva dei parlanti. Questa scelta metodologica ha il vantaggio di restituire in maniera più trasparente la competenza del parlante, potendo poi rapportare i risultati ottenuti dallo studio di *rating* ai dati della frequenza del corpus.

La tesi è stata organizzata come segue. Nel primo capitolo sono stati descritti i principali modelli di recupero lessicale in psicolinguistica. Nel secondo capitolo invece viene descritto lo studio quantitativo condotto sul corpus e sul *rating*. Infine, nel terzo capitolo viene avanzata una proposta generale del fenomeno della *plural dominance* e sono esposte le discussioni finali.

Capitolo 1. Principali modelli del recupero lessicale in psicolinguistica

1.1. Modelli in comprensione.

1.1.2 Modello Parallel Dual-Route

Particolare attenzione è stata rivolta verso la formulazione di modelli psicolinguistici riguardanti il recupero lessicale che potessero rendere conto anche del fenomeno della *plural dominance*. In generale, in studi quali Baayen et al. (1996; 1997) si sono riscontrati tempi di reazione più lunghi per i singolari a dominanza plurale rispetto agli stessi nella condizione di *singular dominance*. Pertanto, lo studio delle parole di tipo *plural dominant* rappresenta un banco di prova per quanto riguarda la potenza predittiva dei modelli di recupero lessicale. Infatti, come chiarito in quanto segue, non tutti i modelli di recupero lessicale rendono conto in maniera efficace di questo fenomeno.

In comprensione, un primo modello teorico sul riconoscimento di parole morfologicamente complesse è quello proposto da Butterworth (1983), il quale prevede un'entrata di tipo *full-list* per tutte le parole. Questo modello prevede quindi che i termini vengano elaborati nella loro forma intera, senza essere scomposti in morfemi. Secondo un'ipotesi lessicale di questo genere, le parole dovrebbero essere sensibili alla frequenza di occorrenza della parola intera, non prendendo in considerazione in alcun modo la frequenza cumulativa della radice. Ciò comporterebbe dei tempi di risposta minori per le forme singolari dei nomi *singular dominant*, essendo quest'ultimi i più frequenti per occorrenza di parola, mentre i singolari dei nomi *plural dominant* dovrebbero elicitarne le risposte più lente. Invece, alcuni studi sperimentali hanno evidenziato come forme singolari di tipo *plural dominant* elicitino dei tempi di risposta più brevi (Baayen et al., 1996; 1997).

Concettualmente opposto e' invece l'approccio noto come *full-parsing*, a decomposizione obbligatoria, di cui e' un esempio Taft (1975). Questo modello prevede un'operazione chiamata *affix stripping* secondo la quale per poter riconoscere una parola morfologicamente complessa in prima istanza è necessario decomporre la radice dall'affisso. In questo caso, il tempo di recupero lessicale sarebbe in assoluto più lento rispetto a quello previsto nei modelli di tipo *full-listing* poiché la mente non recupererebbe qualcosa di già completo, ma dovrebbe costruire la

parola a partire dalle varie combinazioni di morfemi conformemente alle regole grammaticali. Quindi, da un punto di vista empirico, si prevederebbero dei tempi di risposta simili per le forme singolari e quelle plurali di un nome, condividendo entrambi la stessa frequenza cumulativa di radice. Tuttavia, alcuni lavori in letteratura sembrano suggerire che, in generale, le forme plurali richiedono tempi di risposta maggiori rispetto alle forme singolari (vedi ad es. Sereno e Jongman, 1997).

Un terzo tipo di modelli, chiamati *Dual-Route*, congiungono i punti di forza dei primi due: in questo caso, infatti, a seconda del termine presentato, si può attivare o la via diretta che permette di recuperare la parola nella sua forma intera (e, come abbiamo notato in precedenza, più rapida), oppure la via decomposizionale che permette di recuperare una parola a partire dalla radice più gli affissi. Un esempio di modello di tipo *Dual-Route* è l'AMM (*Augmented Addressed Morphology*) proposto da Caramazza (1988), in cui la via decomposizionale e la via diretta si attivano in modo complementare: la prima sarebbe attiva solo nel recupero di parole a bassa frequenza, neologismi e pseudo-parole, mentre la seconda sarebbe attiva solo nel recupero di termini già noti al parlante.

Il modello noto come *Parallel Dual-Route* proposto da Schreuder e Baayen (1995) è anch'esso un modello a doppia via, ma si distacca rispetto agli altri poiché prevede l'attivazione contemporanea sia della via diretta (*full-listing*) che della via decomposizionale (*full-parsing*). Nel

modello AMM, invece, la via decomposizionale si attiverebbe solo nel caso in cui la via diretta avesse fallito il recupero del nome. Il grande vantaggio del modello *Parallel Dual-Route* consiste quindi nel fatto che le due vie possono lavorare simultaneamente, consentendo in questa maniera un'elaborazione più rapida. L'accesso all'una o all'altra via dipende però sia dalla frequenza della radice e sia dalla frequenza della parola intera: in genere, parole a bassa frequenza saranno elaborate con maggiore probabilità attraverso la via decomposizionale, mentre termini a più alta



Figura 1:
Rappresentazione
Del modello
Parallel Dual-
Route

frequenza saranno elaborati per mezzo della via diretta. Secondo questa ipotesi, dunque, nel caso dei plurali di nomi *singular dominant* è più probabile che per essi si verifichi l'attivazione della via decomposizionale, mentre per i plurali di nomi *plural dominant* viene privilegiata la via diretta, essendo quest'ultima sensibile all'alta frequenza della parola intera. Per i nomi singolari, invece, a determinare la velocità con la quale vengono riconosciuti un ruolo importante lo gioca la frequenza assoluta del lessema. Le forme singolari, infatti, vengono facilitate dall'alta frequenza delle forme plurali, mandando quest'ultima attivazione retroattiva per un riconoscimento più rapido dei singolari.

1.1.3 Baayen, Dijkstra, e Schreuder, (1997)

I risultati degli esperimenti sull'olandese descritti in Baayen et al. (1997) sono compatibili con le predizioni del modello *Parallel Dual-Route*. Il primo di questi esperimenti, un compito di decisione lessicale, richiedeva la somministrazione di quattro batterie di nomi differenti rispetto a due variabili, cioè la frequenza della radice e la frequenza della parola intera. Più in dettaglio, la prima condizione prevedeva nomi ad alta frequenza di radice di tipo *singular dominant*; nella seconda condizione i nomi erano ad alta frequenza di radice, ma di tipo *plural dominant*; nella terza condizione erano presenti nomi a bassa frequenza di radice di tipo *singular dominant* ed infine la quarta condizione comprendeva nomi a bassa frequenza di radice di tipo *plural dominant*.

I risultati dell'esperimento mostrano che gli stimoli costituiti da forme al singolare, di nomi sia di tipo *singular dominant* che di tipo *plural dominant*, quando condividono una frequenza della radice comparabile, elicitano tempi di risposta simili, mentre le forme plurali mostrano un effetto di frequenza per l'occorrenza della parola. Infatti, le forme plurali di nomi *plural dominant* sono associate tempi di reazione minori rispetto alle forme plurali dei nomi *singular dominant*. Il vantaggio associato alle forme plurali dei nomi *plural dominant* sarebbe dovuto presumibilmente ad una sovrapposizione della via diretta su quella decomposizionale. Il modello matematico ideato da Baayen et al. (1997) pone infatti in evidenza come la via decomposizionale sia molto costosa e sia più efficace e veloce della via diretta solamente nel caso di parole plurali a bassa frequenza di radice.

Nel secondo esperimento, invece, le forme singolari non sono state pareggiate per la frequenza della radice, ma unicamente per l'occorrenza della parola intera. Si sono presentate

tre condizioni: nella prima la frequenza delle forme singolari e plurali dei nomi somministrati era la stessa; nella seconda condizione gli stimoli presentavano una dominanza per le forme singolari; mentre nella terza condizione i nomi erano di tipo *plural dominant*. I risultati hanno mostrato che per le forme singolari quando è più alta la frequenza dei corrispettivi plurali, minore è il loro tempo di risposta. Questo compito di decisione lessicale mostra dunque che quando la frequenza della parola intera del singolare è mantenuta costante, il tempo di reazione varia significativamente come una funzione dei plurali.

Dallo studio di Baayen et al. (1997) emerge dunque l'importanza per un modello di recupero lessicale di tenere conto della variabile frequenza. La sensibilità alla frequenza, infatti, descrive al meglio i risultati ottenuti nel suo esperimento, laddove invece gli approcci di tipo *full-listing* e *affix-stripping* falliscono. Un modello come quello proposto da Butterworth (1983) avrebbe predetto infatti dei tempi più lunghi per i singolari a dominanza plurale, non prevedendo che essi potessero essere sensibili alla frequenza cumulativa della radice. Secondo un'ipotesi a decomposizione obbligatoria, invece, nomi di tipo *singular dominant* e *plural dominant* avrebbero dovuto elicitare dei tempi simili. Solo con un modello a doppia entrata invece si spiegano gli effetti della frequenza nell'attribuire un vantaggio ai nomi di tipo *plural dominant*.

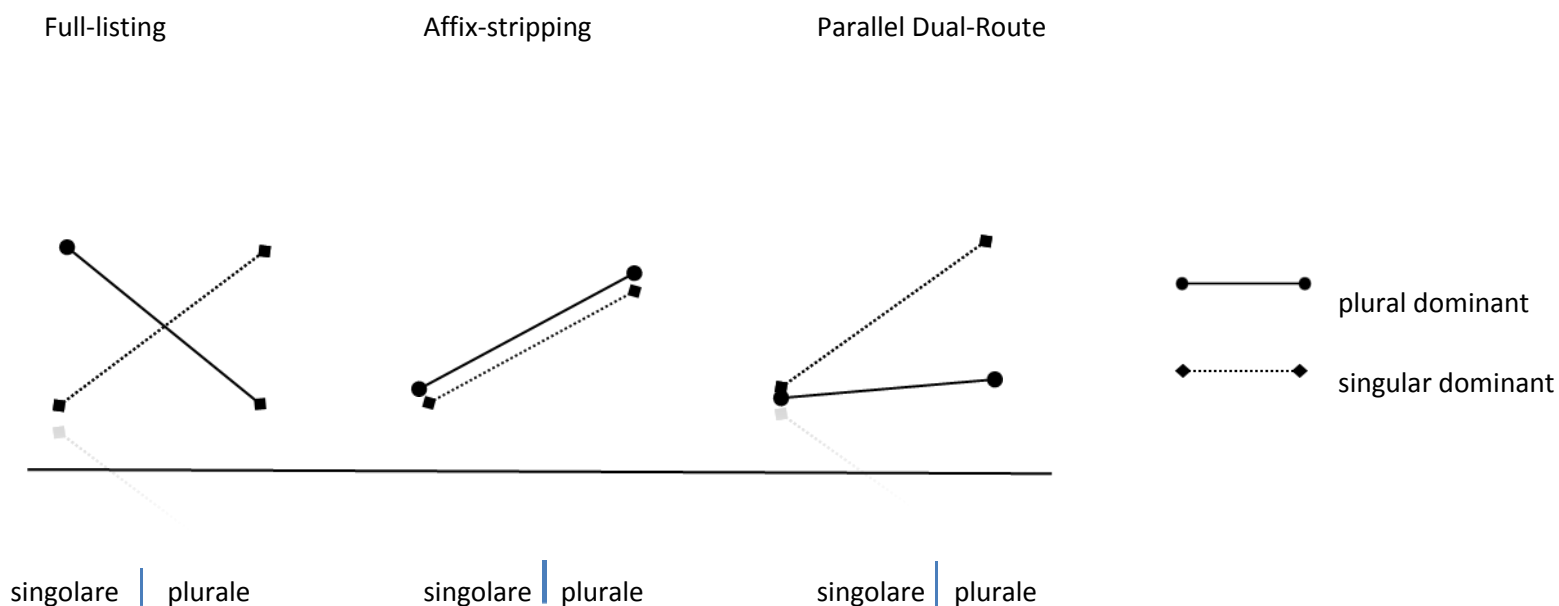


Figura 2: Tempi di reazione dei nomi *singular* e *plural dominant* per i diversi modelli

1.1.4 Baayen, Burani, e Schreuder, (1996)

Gli stessi esperimenti di decisione lessicale in olandese sono stati condotti da Baayen e colleghi (1996) anche in lingua italiana. In italiano, rispetto all'inglese o all'olandese, anche le forme singolari sono parole morfologicamente complesse. Questo non è un fatto secondario: infatti, in italiano è possibile neutralizzare la differenza tra forma singolare e forma plurale di un nome rispetto alla lunghezza ortografica e di conseguenza i risultati sperimentali sono più indicativi.

Il primo esperimento svolto in questo studio è metodologicamente uguale al compito di decisione lessicale descritto in Baayen et al. (1997). I risultati per questo esperimento mostrano che in italiano le forme singolari dei nomi di tipo *singular dominant* e *plural dominant* vengono recuperati quasi con la stessa velocità, suggerendo perciò che i singolari non vengano elaborati per la via diretta, ma decomposizionalmente secondo la formula base+suffisso. Apparentemente dunque, come per l'olandese, anche in lingua italiana i tempi di reazione delle forme singolari sono determinati dalla frequenza cumulativa della radice.

Per le forme plurali, invece, si osserva un forte effetto della frequenza di occorrenza della parola intera. Nonostante infatti i plurali siano stati pareggiati per frequenza della radice, i nomi del tipo *plural dominant* sono stati elaborati con molta più rapidità di quelli a dominanza singolare, suggerendo quindi che per le forme plurali dei nomi di tipo *plural dominant* venga attivata la via diretta. Per le forme plurali dei nomi di tipo *singular dominant* ad entrare in azione sarebbe invece la via decomposizionale.

Comparando i risultati con quelli effettuati da Baayen et al (1997) in lingua olandese, si notano nei risultati in lingua italiana dei tempi di risposta più brevi per le forme plurali dei nomi di tipo *plural dominant*, le quali in assoluto sono state le forme elaborate in maniera più rapida.

Questo studio conferma i risultati che abbiamo descritto in Baayen et al. (1997), enfatizzando la potenza predittiva del modello *Parallel Dual Route* come quadro esplicativo per l'interpretazione dei dati.

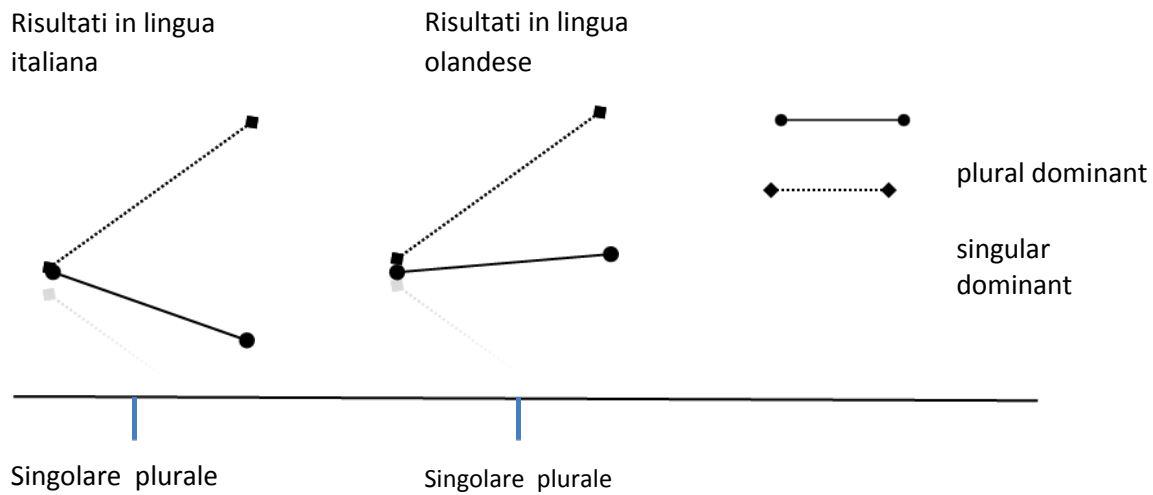


Figura 3: tempi di reazione dei compiti di decisione lessicale in Baayen et al. (1996) e Baayen et al. (1997)

1.2. Modelli in produzione

1.2.1 Levelt, Roelofs, e Mayer, (1999)

Un modello classico per il recupero del lessico in produzione è quello spiegato in Levelt et al. (1999). Semplificando, un concetto lessicale attiva alcuni nodi lemmatici, ma solo il lemma con maggiore attivazione viene selezionato. In questo modello, i lemmi vengono recuperati in forma intera, e, col recupero, diventano disponibili anche le proprietà sintattiche come genere, numero, tempo, persona, ecc. Una volta selezionata la sintassi del lemma, l'informazione procede verso lo stadio fonologico/articolatorio. Il passo successivo consiste nel recuperare la forma fonologica del lemma dal lessico mentale (il lessema). L'accesso al lessema è uno stadio del processo di particolare interesse poiché è il luogo in cui incide la frequenza (vedi Jeschniak e Levelt, 1999). La parola viene infine articolata.

In questo studio, il fenomeno della *plural dominance* assume per la prima volta un ruolo di rilievo poiché viene proposto un modello rappresentativo per l'elaborazione lessicale in fase di produzione specifico anche per i nomi di tipo *plural dominant*. In particolare, il numero viene rappresentato lungo i livelli appena descritti. Nel primo livello, quello semantico, viene assegnata la proprietà "*multiple*", la quale indica un numero di esemplari di un medesimo elemento maggiore di uno. La proprietà "*multiple*" quando innescata, attiva al livello del lemma (livello grammaticale) una serie di tratti, quali ad esempio "plurale" o "numerabile". Per ultimo queste informazioni arrivano al livello della forma di parola (lessema) dove si compongono i suffissi

plurali.

Per quanto riguarda la differenza fra parole di tipo *singular* e *plural dominant*, Levelt et al.(1999) propongono differenti rappresentazioni. I nomi *plural dominant*, diversamente dai nomi di tipo *singular dominant*, hanno due rappresentazioni diverse, una per la forma singolare e un'altra per la forma plurale, già al livello semantico e a quello del lemma. Tuttavia, al livello del lessema, sia i nomi di tipo *singular dominant* che quelli di tipo *plural dominant* convergono verso lo stesso nodo per comporre la forma plurale in maniera decomposizionale. A distinguere le due condizioni di dominanza sarebbero dunque la differente concettualizzazione semantica e lo stadio sintattico, mentre al livello del lessema non vi sarebbero distinzioni.

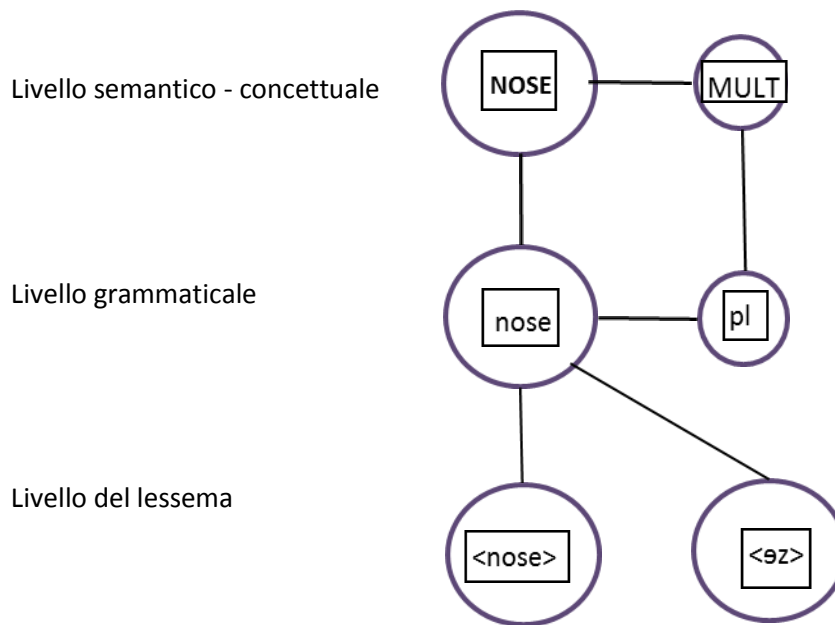


Figura 4: Rappresentazione del recupero lessicale in produzione dei nomi *singular dominant* secondo Levelt et al. (1999)

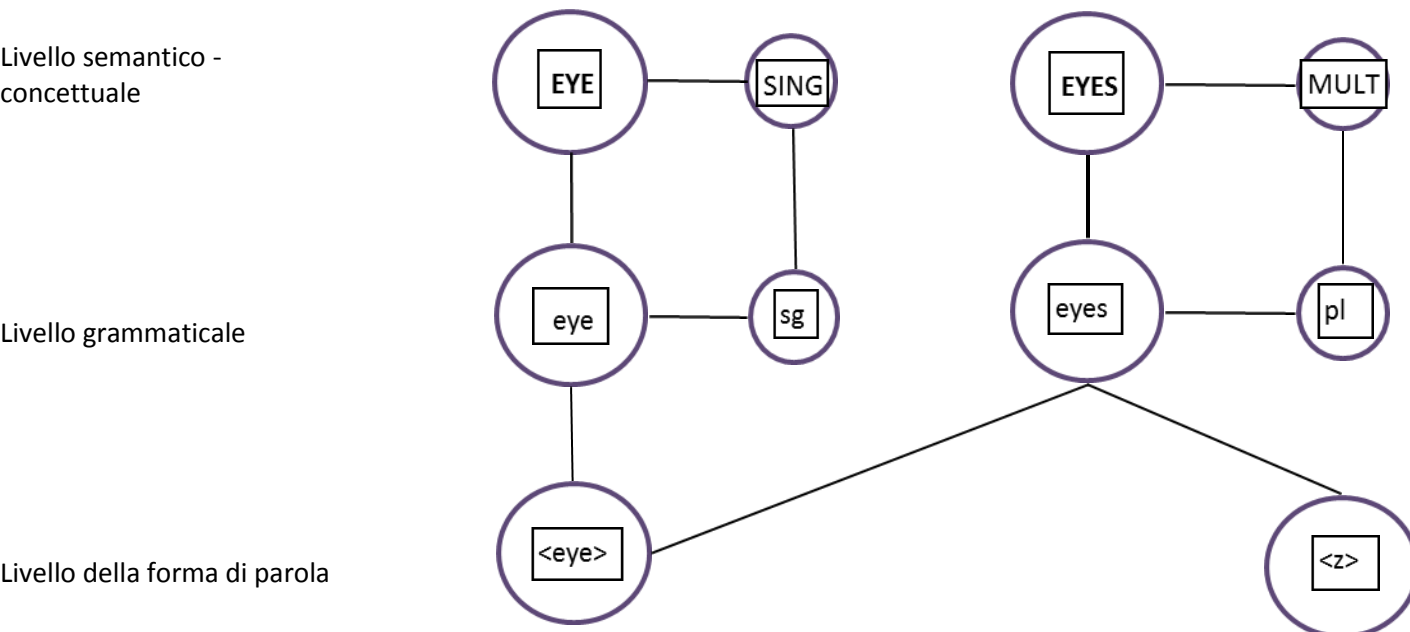


Figura 5: Rappresentazione del recupero lessicale in produzione dei nomi *plural dominant* secondo Levelt et al. (1999)

Per testare la produzione linguistica, un tipo di esperimento cui maggiormente si ricorre in letteratura è la denominazione di figure. Un esempio, sempre rispetto al fenomeno della *plural dominance*, è rappresentato dal lavoro di Baayen, Levelt e Haveman discusso in Levelt et al. (1999).

I soggetti partecipanti all'esperimento vedevano delle immagini raffiguranti un singolo oggetto o il medesimo oggetto reiterato più volte; i soggetti dovevano denominare l'immagine usando la forma plurale o la forma singolare dell'oggetto/degli oggetti rappresentati. I nomi degli oggetti erano sia di tipo *singular dominant*, sia di tipo *plural dominant*, e gli stimoli sono stati pareggiati per la frequenza cumulativa della radice.

I risultati hanno mostrato un tempo di latenza più lungo per i plurali rispetto ai singolari, dato aspettato essendo i plurali in inglese morfologicamente più complessi. Sorprendentemente però, entrambe le forme singolari e plurali dei nomi di tipo *plural dominant* sono state più lente di quelle a dominanza singolare, particolarmente nel caso dei plurali *plural dominant* che sono stati elaborati nel tempo più lungo, in maniera opposta a quanto abbiamo visto verificarsi per i compiti di decisione lessicale (Baayen et al., 1997). Non vi è stato dunque nessun effetto per la frequenza di occorrenza della parola intera, fenomeno questo dovuto secondo la spiegazione di Levelt et al. (1999) al fatto che i lemmi singolari e plurali dei nomi di tipo *plural dominant* condividono una stessa entrata al livello del lessema. In questo punto le forme plurali infatti vengono composte aggiungendo i suffissi alla base, ed essendo questa un'operazione che non

procede attraverso la via diretta ad influire sarebbe quindi la sola frequenza cumulativa della radice.

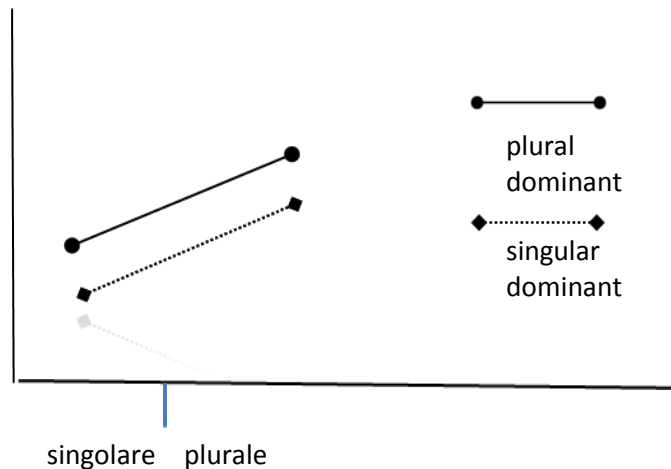


Figura 6: Tempi di risposta per nomi di tipo *singular* e *plural dominant* nel compito di denominazione di figure in Levelt et al. (1999)

Ipotizzare nel modello di elaborazione lessicale due entrate a livello di concetto e di lemma per i nomi *plural dominant* spiegherebbe il ritardo riscontrato nel compito di denominazione di figure dei termini *plural dominant* rispetto alle parole a dominanza singolare. Una doppia entrata concettuale creerebbe una competizione nella selezione lessicale e causerebbe dei tempi di risposta più lunghi. Il fatto che vi siano due concetti accoglie l'osservazione di Booij (1996) per cui una pluralizzazione non sarebbe semplicemente un'operazione di moltiplicazione, ma si tratterebbe di una sorta di derivazione che farebbe assumere alla forma plurale di una parola una denotazione diversa. Questa derivazione renderebbe quindi la pluralizzazione una forma concettualmente e semanticamente diversa rispetto alla corrispettiva forma singolare. Più in dettaglio, ciò sembra essere particolarmente applicabile alle forme plurali dei nomi di tipo *plural dominant*: quest'ultime infatti, comparando più frequentemente al plurale, sono più soggette al processo di lessicalizzazione secondo il quale acquisirebbero una distinta idiosincronicità semantica.

1.2.2 Baayen, Levelt, Schreuder, e Ernestus, (2007)

Più recenti esperimenti in fase di produzione sono descritti in Baayen et al. (2007). In un primo esperimento di denominazione di figure si sono scelte immagini variate per alta/bassa frequenza. I nomi *plural dominant* hanno elicitato anche in questo caso tempi di reazione più alti dei nomi *singular dominant*, sia per le forme singolari che per quelle plurali. Ricordiamo che, nell'esperimento di comprensione di Baayen et al. (1997), i plurali a dominanza plurale erano recuperati più velocemente rispetto ai plurali dei nomi di tipo *singular dominant*. In produzione, invece, sono i nomi *plural dominant*, tanto al singolare quanto al plurale, a elicitare tempi di reazione più lenti rispetto ai nomi *singular dominant*. Ovvero, in questo caso, la frequenza di occorrenza della parola intera non sembra giocare alcun ruolo.

Come abbiamo precedentemente analizzato, in Levelt et al. (1999) questa differenziazione nei tempi di reazione era stata argomentata ipotizzando che per i soli nomi *plural dominant* fosse presente per i plurali un'interpretazione semantica differente, e quindi, di conseguenza, un ulteriore lemma di rappresentazione. La presenza di una doppia entrata, una per il singolare e una per il plurale, motiverebbe dunque una sorta di competizione durante il processo di selezione lessicale che causerebbe il ritardo nei tempi di reazione.

Baayen et al. (2007) nel loro lavoro interpretano lo svantaggio dei nomi *plural dominant* sui *singular dominant* in modo totalmente diverso, secondo un approccio più prettamente statistico. In questo studio, infatti, il concetto di dominanza viene sostituito con quello di entropia (flessiva), concetto introdotto dalla teoria dell'informazione (Shannon & Weaver 1949). In questo caso, l'entropia flessiva misura la probabilità dato un lessema di presentarsi nella forma singolare o plurale rispetto al proprio paradigma. I nomi a dominanza plurale presentano un grado di entropia flessiva più alta rispetto ai nomi a dominanza singolare in quanto in genere le proprie forme plurali non eccedono di molto per frequenza le corrispettive forme singolari. Nei nomi di tipo *singular dominant*, invece, mediamente la forma singolare è significativamente più frequente della propria forma plurale, riflettendo pertanto un grado di entropia flessiva minore.

Il vantaggio nel prendere in considerazione questa nozione è quello di rendere conto nel fenomeno della *plural dominance* della gradualità dei nomi per la variabile frequenza, senza dicotomizzare le parole in *singular* e *plural dominant*. In questo modo, pertanto, non si ha necessità di dover spiegare la maniera in cui una forma plurale di tipo *plural dominant* acquisirebbe una seconda entrata del lemma, poiché di per sé non si ipotizzano nomi a

dominanza singolare e plurale come due categorie ben definite.

Inoltre, altro dato essenziale da introdurre per poter fare delle analisi qualitative, è per Baayen et al. (2007) anche quello di entropia relativa che quantifica il distacco per cui la probabilità di distribuzione di un particolare nome diverge dalla corrispondente probabilità di distribuzione della sua classe di nomi. Relazionandolo al concetto di dominanza, quindi, più alto è il grado di entropia relativa e più aumenta la probabilità che il nome sia di tipo *plural dominant* poiché è più grande la sua divergenza da quella che è la normale distribuzione di occorrenza della flessione, rappresentata dalla condizione dei nomi di tipo *singular dominant*.

Rimane tuttavia senza spiegazione, sia in Baayen et al. (2007) quanto in Levelt (1999), un dato fondamentale. Come si spiega la divergenza tra i risultati ottenuti in produzione e quelli ottenuti in comprensione, ad esempio in Baayen et al. (1997)? Per quale motivo variando il compito dalla decisione lessicale alla denominazione di immagini i due meccanismi non mostrano dei *pattern* di risposta simili? Baayen e colleghi (2007) provano a fornirne una spiegazione, in un passaggio infatti si afferma (p. 23, Baayen et al., 2007):

" When we consider the production and comprehension data jointly, an important similarity is the presence of a facilitatory lexeme frequency effect. This suggests that in both modalities an abstract representation for the lexeme (in the sense of Aronoff, 1994, or a lemma in the sense of Levelt, 1989) is accessed first. What is different between the two modalities is the subsequent access to the inflected forms given the lexeme. In comprehension, it is the probability of the singular (or plural) form given the lexeme that comes into play. The greater an inflected form's paradigmatic probability, the faster it is accessed. In production, it is the entropy of the paradigm that is at issue, at least for singulars ".

Le differenze riscontrate fra le due modalità, dunque, vengono motivate da Baayen et al. (2007) a partire da alcune diversità nell'accesso alla forma flessa del lemma: in comprensione è la nozione di entropia flessiva che gioca un ruolo maggiore, mentre in produzione ha un peso maggiore l'entropia relativa. Ossia, si ritiene che i nomi di tipo *plural dominant* abbiano dei paradigmi a più alta informazione rispetto ai nomi *singular dominant* e questo sembrerebbe causare dei costi nell'accesso in produzione, mentre li faciliterebbe in comprensione.

L'aspetto più importante di questo studio, comunque, è quello di avere dato rilievo nelle

analisi dell'elaborazione del numero morfologico anche alla forza del paradigma del nome. Un'analisi basata solo sulle regole di derivazione dalla radice, infatti, sottostimerebbe la complessità dell'organizzazione del lessico mentale rispetto alle sue relazioni paradigmatiche. Inoltre, prendendo in considerazione l'intero paradigma, lo status speciale dei nomi a dominanza plurale in produzione potrebbe derivare dai costi di accesso al paradigma del lessema, e conseguentemente questo potrebbe indicare che il fenomeno della *plural dominance* non si origini allo stadio del lemma come sostenuto da Levelt et al. (1999), bensì a uno stadio successivo e più formale concernente appunto il paradigma morfologico del lessema.

1.2.3 Un approccio diverso: Biedermann, Beyersmann, Mason, e Nickels, (2013)

In anni recenti Biedermann in collaborazione con altri studiosi ha messo in atto alcuni esperimenti per fare più chiarezza sull'argomento del recupero lessicale dei nomi *plural dominant* in fase di produzione. La motivazione nasce dal fatto che se vogliamo asserire che i nomi a dominanza plurale abbiano uno statuto speciale nel lessico è necessario osservare per essi degli effetti specifici sia in produzione sia in comprensione.

In Biedermann et al. (2013) vengono somministrati alcuni compiti di denominazione di immagini (che testano la produzione) in lingua inglese sia a soggetti sani che a soggetti afasici non fluenti con lo scopo di cercare dei risultati che siano coerenti con i fenomeni riscontrati anche in comprensione. In questi esperimenti, le forme singolari dei nomi *plural dominant* possiedono sempre una frequenza della parola intera più bassa rispetto alle forme singolari dei nomi *singular dominant*; la frequenza di radice cumulativa è invece invariata tra i due gruppi di nomi. Le forme plurali dei nomi *singular* e *plural dominant*, invece, sono pareggiate per frequenza di occorrenza della parola intera: questo dato è di notevole importanza perché se si volesse relegare l'intero fenomeno della *plural dominance* ad una questione di mera frequenza, i tempi di reazione elicitati dalle forme plurali di entrambe le tipologie di nomi dovrebbero essere simili.

Il primo compito di denominazione di immagini è stato presentato ai soggetti sani i quali hanno elaborato più velocemente i singolari dei nomi *singular dominant* rispetto agli stessi dei nomi *plural dominant*. Sorprendentemente rispetto a quanto abbiamo osservato finora, nella condizione di *plural dominance* le forme singolari e le forme plurali sono state elaborate in un

tempo simile. Le forme plurali dei nomi *singular dominant* invece hanno elicitato i tempi di risposta più lunghi. Non è stato rilevato, perciò, nessuno svantaggio per le forme plurali dei nomi *plural dominant* rispetto alle forme plurali dei nomi *singular dominant*, contrariamente a quanto osservato invece in Baayen et al. (2007) e Levelt et al. (1999).

Senza dubbio, si tratta di risultati cruciali rispetto al quadro presentato fino ad ora. Non avendo rilevato un ritardo per i nomi *plural dominant* in produzione Biedermann et al. (2013) ipotizzano infatti che non vi sia una diversificazione semantica fra le due classi (come avveniva invece in Levelt et al., 1999), ma nomi di tipo *singular* e *plural dominant* divergerebbero solo al livello lessicale dove incide anche la variabile della frequenza. Quello che appare chiaro dai dati rilevati in Biedermann et al. (2013) è che questi risultati trovano spiegazione solo alla luce del modello *Parallel Dual-Route* descritto in precedenza (paragrafo 1.2.1). E' solo questo modello, infatti, che ipotizza i plurali dei nomi *plural dominant* come entrate lessicali.

Biedermann et al. (2013) hanno indagato il fenomeno della *plural dominance* anche in afasia, attraverso un compito di denominazione di immagini. All'esperimento si sono sottoposti SJS e DEH, due pazienti afasici non fluenti. Gli esiti hanno dimostrato che entrambi i pazienti hanno compiuto molti errori di numero. SJS ha avuto più problemi con il plurale in generale, mentre in condizione di *singular dominance* i plurali sono stati problematici per entrambi i soggetti. Nella condizione di *plural dominance*, invece, nessuno dei due soggetti ha mostrato una differenza significativa nella produzione fra forma singolare e forma plurale.

Questi risultati, dunque, concordano con gli studi effettuati in comprensione (Baayen et al., 1996; 1997), ma sono in contrasto, come abbiamo constatato, con quanto rilevato in Baayen et al. (2007): in questo studio le forme plurali dei nomi di tipo *plural dominant* elicitavano risposte più lunghe. Inoltre, anche in afasia è stato dimostrato che gli effetti di dominanza non appaiono al livello concettuale o del lemma. Nemmeno in fase post-lessicale si sono osservate differenze, per cui ciò sembra provare che il fenomeno della dominanza riguardi il livello fonologico della forma di parola o il suo accesso.

Sulle discrepanze che hanno portato a due risultati così diversi, si afferma in Biedermann et al. (2013) che non sembra possibile siano il frutto del grado della frequenza relativa degli stimoli. Baayen e colleghi avevano utilizzato set di parole a bassa ed alta frequenza, mentre nello studio appena presentato si sono utilizzati principalmente nomi a più bassa frequenza, ma, comparando i risultati di Baayen et al. (2007) considerando solo gli stimoli a bassa frequenza si nota che sono di fatto simili a quelli di Biedermann et al. (2013). Una possibile causa potrebbe derivare dalla differenza fra le due lingue utilizzate per gli studi (inglese in Biedermann et al.,

2013; e olandese in Baayen et al., 2007). Nomi a dominanza plurale non corrispondono necessariamente nelle due lingue e in olandese ad esempio il sistema morfologico è più complesso, mostrando cinque diverse uscite flessive per i nomi plurali contro la sola regolare in -s per l'inglese. Inoltre, l'aggiunta di morfologia plurale in olandese in molti casi causa un allungamento della sillaba, mentre questo fenomeno avviene solo raramente in lingua inglese (ad es. nel sg. *horse* e pl *horse-s*). Di fatto però, i motivi di una tale divergenza nei risultati non sono ancora stati individuati e ulteriori ricerche sono necessarie, anche in lingue tipologicamente diverse, per avere un'immagine più nitida del fenomeno.

1.2.4 Luzzati, Mondini, e Semenza, (2001)

Un altro studio compatibile con il modello *Parallel Dual Route* è quello condotto su persone afasiche di lingua italiana da Luzzati et al. (2001). Più in dettaglio, si è lavorato con un soggetto affetto da dislessia fonologica e agrammatismo, la cui via di conversione grafema-fonema era compromessa (in questi casi si riscontrano problemi con pseudo-parole e parole nuove, mentre parole ad alta frequenza sono conservate).

Con una prova in particolare si mirava ad accertare eventuali differenze nella produzione fra nomi regolari e irregolari, nomi *singular* o *plural dominant* o per parole contenenti dei suffissi valutativi. In particolare, da questa prova è emerso che le forme singolari dei nomi *singular dominant* sono state articolate in modo significativamente più accurato rispetto alle relative forme plurali. La produzione dei nomi *plural dominant*, invece, si colloca all'interno di questi due poli e non viene riportata alcuna differenza significativa fra le forme al singolare e le corrispettive plurali.

L'analisi qualitativa dei risultati ha evidenziato che nella condizione di *singular dominance* il paziente non ha mai sostituito una forma singolare con una forma plurale: le sostituzioni sono infatti avvenute solo nella direzione opposta. Per quanto riguarda i nomi *plural dominant*, invece, gli errori di sostituzione sono stati commessi in entrambe le direzioni e con frequenza equiparabile. Anche questi dati, pertanto, supportano il fatto che le parole ad alta frequenza relativa vengano analizzate come unità non decomposte e che, in particolare, le forme plurali dei nomi *plural dominant* siano meno complesse per quanto riguarda la loro elaborazione se comparate alle forme plurali dei *singular dominant* grazie al loro maggiore

numero di occorrenze rispetto alla base.

1.2.5 Biedermann, Lorenz, Beyersmann, e Nickels, (2012)

Uno studio che rileva dei dati differenti ancora da quelli presentati fino ad ora è quello di Biedermann et al. (2012). In questo lavoro viene esaminato in particolare un paziente afasico fluente chiamato DRS. DRS presenta problemi di natura semantico-lessicali, mentre il livello concettuale è preservato. Inoltre, presenta alcuni deficit causati da un danneggiamento al livello del lemma o all'accesso del lessema.

. In un compito di denominazione di immagini DRS ha commesso più errori di numero relativamente alle forme plurali e indipendentemente dalla categoria. Tuttavia, ha manifestato maggiori problemi nel recuperare le forme plurali dei nomi *singular dominant*. DRS era in grado per la maggior parte dei casi di recuperare la radice del nome (il lemma), ma non di aggiungere la morfologia plurale.

Anche nel compito di associazione figure e parole, il paziente ha commesso solo errori di numero. La sua produzione è caratterizzata da un effetto legato al tipo di dominanza. Infatti, forme plurali *singular dominant*, sostituite la maggior parte delle volte con forme al singolare, sono risultate significativamente più compromesse delle forme plurali *plural dominant*. Dei risultati di questo tipo sia in comprensione che in produzione, pertanto, localizzerebbero il deficit al livello del lemma, e di conseguenza anche il fenomeno della *plural dominance*.

. Questi dati sono quindi in parte compatibili con il modello di Levelt et al. (1999) in cui il lemma gioca un ruolo fondamentale nella differenziazione dei nomi *plural dominant* da quelli a *singular dominance*. Tuttavia, essendo le forme dei plurali *plural dominant* recuperate con meno problematicità rispetto a quelle a dominanza singolare, differentemente da Levelt et al. (1999) non si può ipotizzare che forme singolari e plurali dei nomi di tipo *plural dominant* condividano un unico nodo al livello del lessema con i nomi di tipo *singular dominant*. I risultati inoltre non trovano completa spiegazione nemmeno adottando la proposta di Biedermann et al. (2013) in cui si prevede un effetto di dominanza al solo livello della forma di parola, non ipotizzando nessuna sensibilità per la frequenza al livello grammaticale.

1.2.6 Nickels, Biedermann, Fieder, e Schiller, (2015)

Un ulteriore modello empirico rappresentante la formazione lessicale in interazione col numero che prendiamo in esame è quello in Nickels et al. (2015).

In questo studio viene discusso il modello di Levelt per il recupero lessicale in produzione precedentemente esposto nel paragrafo 1.2.1 e qui riproposto nella figura N. 7.

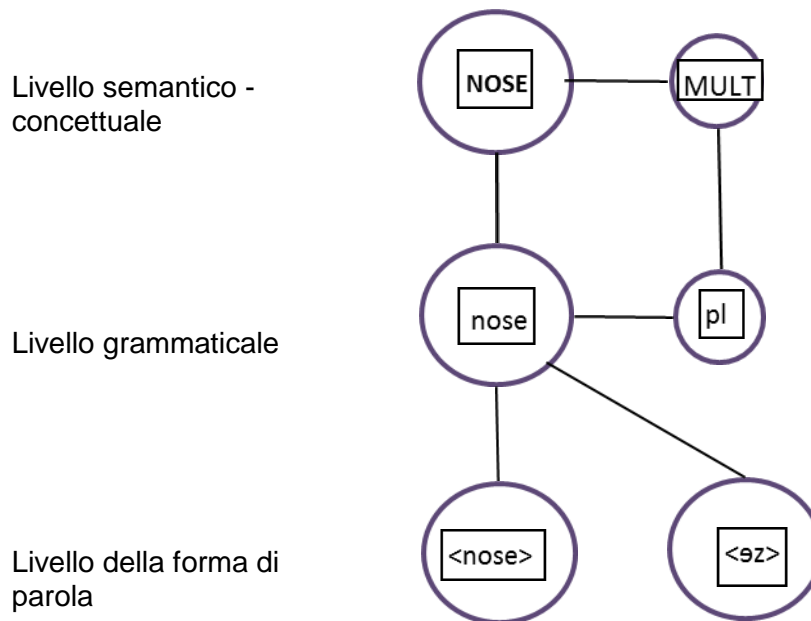


Figura 7: Rappresentazione del recupero lessicale in produzione dei nomi *singular dominant* secondo Levelt et al. (1999)

Riguardo a tale rappresentazione dell'elaborazione lessicale, Nickels e colleghi contestano il fatto che esso non possa predire la forme irregolari dei nomi plurali o le ipergeneralizzazioni che si verificano quando in fase di acquisizione della lingua i parlanti regolarizzano anche le forme irregolari o le non parole. L'alternativa fornita da Nickels et al. (2015) è quella di ipotizzare per i nomi (anche *singular dominant*) due lemmi distinti per le forme singolari e plurali, e, diversamente da quanto abbiamo già visto in precedenza in altri modelli, i due lemmi verrebbero attivati da un singolo concetto. Un modello siffatto predirebbe quindi le ipergeneralizzazioni (punto debole del modello di Levelt), includendo anche le non-parole. In

caso di non-parole, infatti, il nodo "*multiple*" verrebbe ugualmente innescato attivando a sua volta il lemma e il lessema plurale, senza dover passare attraverso il lemma della forma singolare come avveniva precedentemente in Levelt et al. (1999).

L'idea della presenza di un doppio lemma, come abbiamo precedentemente analizzato, era già stata avanzata da Levelt et al. (1999), in cui si proponeva di separare i lemmi per il singolare e il plurale nel solo caso però dei nomi a *plural dominance*. La ragione di questa separazione era stata spiegata concettualmente facendo riferimento alla differenza semantica che esiste per certi nomi plurali rispetto ai loro singolari, rimarcando dunque che in questi termini la pluralità non indica solamente il multiplo di un'entità, ma avrebbe anche delle ripercussioni sul piano semantico. Pertanto, secondo Levelt et al. (1999), nei nomi *plural dominant* non vi sarebbe solamente un secondo lemma, ma anche un concetto differente, separato da quello dei nomi a dominanza singolare. Questo determinerebbe uno svantaggio in produzione per i nomi *plural dominant* i quali pagherebbero durante la fase di selezione lessicale i costi generati dalla presenza di una doppia entrata a livello concettuale e di lemma.

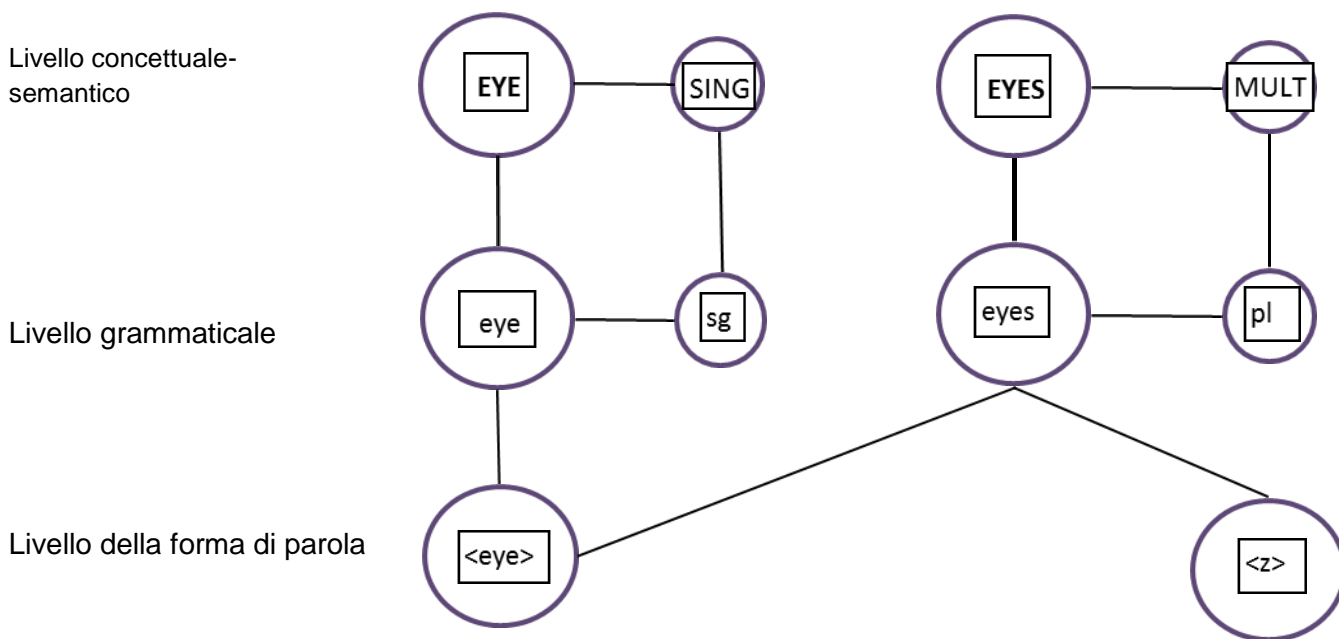


Figura 8: Rappresentazione del recupero lessicale in produzione dei nomi plural dominant secondo Levelt et al. (1999)

Per Nickels, invece, non c'è ragione di introdurre nel modello per le forme plurali anche la presenza di un altro concetto; secondo il suo pensiero, infatti, i plurali condividono con i singolari la stessa rappresentazione concettuale, la pluralizzazione cioè influenza il significato del singolare aggiungendo informazione al numero, ma non ne cambia il senso.

Facendo interagire il modello proposto da Nickels con la nozione di dominanza, essendo il singolare la forma non marcata, questo si riproduce con una maggiore attivazione per il lemma singolare, il quale, qualora vi fosse competizione, verrà selezionato più velocemente. Guardando però ai risultati ottenuti da Biedermann et al. (2012) ci sono delle situazioni in cui il plurale in produzione ottiene dei tempi di reazione più veloci rispetto ai singolari, ovvero nel caso dei nomi *plural dominant*. Come spiegare questi dati nel modello? Semplicemente Nickels et al. suggeriscono che i nomi *plural dominant* abbiano la stessa rappresentazione concettuale e lessico-sintattica a livello di lemma rispetto ai nomi *singular dominant*, ma nei confronti di quest'ultimi differirebbero per il peso di attivazione che intercorre fra lo stadio dei concetti e i nodi del lemma: ossia, il plurale assumerebbe la forma meno marcata del termine. Questa visione è in accordo con Tiersma (1982) in cui si descrivono i nomi *plural dominant* come parole in cui la minore marcatezza passa dal singolare al plurale.

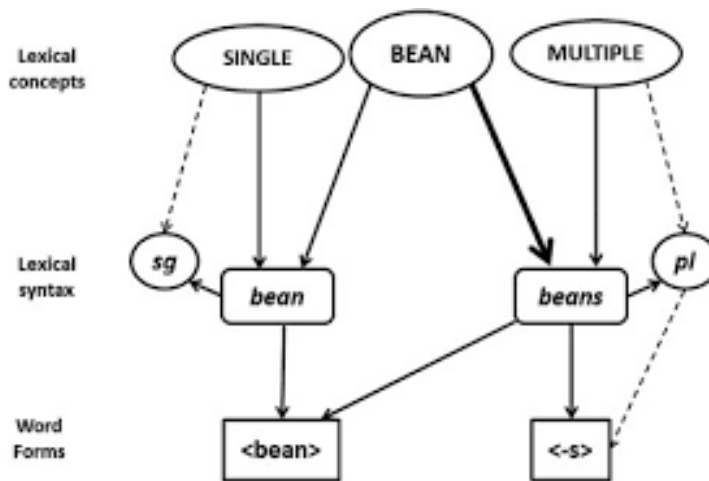


Figura 9: Rappresentazione del recupero lessicale in produzione dei nomi plural dominant secondo Nickels et al.(2015). Le linee tratteggiate indicano attivazioni deboli che permettono le ipergeneralizzazioni, la linea più marcata invece indica il maggiore peso dei plurali come forma più saliente nei nomi plural dominant. In questo modello vi è un unico concetto che attiva due lemmi separati e il diacritico plurale non è più colui che legandosi al nome manda attivazione anche alla forma singolare dal lemma plurale.

1.2.7 Baayen, Milin, Durdevic, Hendrix, e Marelli, (2011)

In un lavoro più recente Baayen e colleghi (Baayen et al. 2011) offrono un modello diverso per spiegare il recupero lessicale di un nome. Il modello viene chiamato *Naive Discriminative Reader* (NDR), e si avvicina alle posizioni connessionistiche (vedi ad es. Harm and Seidenberg, 1999; 2004; Seidenberg and Gonnerman, 2000; Plaut e Gonnerman, 2000). Questo modello è incentrato solo sul livello ortografico e su quello del significato, mentre in comprensione è assente il livello del lessema dove invece un algoritmo stima il peso delle connessioni dalla forma al significato. Mancando il livello del lemma, dunque, viene meno la nozione di morfema, ritenuta inadeguata da un punto di vista teorico per più motivi. Il morfema legato, infatti, può essere usato in contesti diversi possedendo significati diversi e allo stesso tempo in un singolo morfema possono essere contenuti più di un significato. Inoltre, secondo gli autori, non sarebbe sempre facile valutare cosa costituisce un morfema e cosa no (ad esempio gli autori introducono il concetto di fonastema, un digramma che sussiste senza dar vita a una scomposizione morfemica dotata di una precisa nozione semantica (vedi l'elemento *gl-* in *glow*, *gloom*, *glare*, ecc.).

Ne modello NDR i tempi di reazione sono dunque determinati dal modo in cui la rete di associazioni presente nella nostra mente discrimina le forme rispetto ai significati, il tutto sulla base delle frequenza degli elementi che costituiscono il nome. Non esisterebbero pertanto delle rappresentazioni di parole morfologicamente complesse immagazzinate nella mente del parlante, piuttosto la rete di associazioni si affiderebbe alla correlazione tra forme e significati. In questo senso dunque, quanto più una parola segue il proprio paradigma flessivo, tanto prima viene appresa, possedendo una rete di connessioni tra forma e significato più fitta. La frequenza di occorrenza delle parole, quindi, è il fattore determinante nel recupero di una parola dal lessico mentale.

Un vantaggio per questo modello è che forme come i fonastemi trovano una spiegazione potendo nel NDR i digrammi sussistere come elementi indipendenti senza dover ipotizzare per loro una unità come il morfema. La distruzione del concetto del morfema, però, alla stessa maniera assottiglia i confini di modularità presenti tra morfologia e sintassi, rilegando la prima ad essere solo una funzione della seconda. Inoltre, venendo meno l'idea del morfema in un modello di elaborazione lessicale, un'ulteriore ripercussione è che anche il concetto di paradigma flessivo per una categoria lessicale deve essere di conseguenza rivisitato, facendo maggior uso

piuttosto della nozione di densità di vicinato.

Capitolo 2. Uno studio quantitativo: analisi sul corpus e studio di *rating*

In linguistica formale globalmente è stata attribuita poca considerazione al fenomeno della *plural dominance*, maggiore attenzione invece è stata rivolta ad argomenti quali ad esempio i nomi di tipo *mass* e *count*. La ragione di questa divergenza di interesse può essere causata dal fatto che nomi di tipo *mass* tradizionalmente sono collegati a precise classi semantiche e, per quanto riguarda la morfologia di Numero, sembrano mancare della flessione di plurale. I nomi di tipo *plural dominant* invece non sembrano presentare nessun vincolo nè semantico nè morfologico (ad eccezione dei morfemi di singolare presenti in un numero ristretto di nomi e di lingue, i.e. Acquaviva, 2008; Corbett, 2000): è forse questo il motivo della scarsa attenzione ricevuta in linguistica. Tuttavia, è giusto ritenere che anche indagando questo tipo di nomi si possano raggiungere dei risultati interessanti sia per quanto riguarda aspetti dell'elaborazione lessicale (come abbiamo osservato nel capitolo 1), e sia per aspetti concernenti il punto di vista più astratto della linguistica formale. In particolare, studi quantitativi sulla dominanza potranno fornirci dei dati utili per confrontare vari approcci, nel tentativo di comprendere quali siano le variabili, linguistiche e non, che entrano in gioco nell'assegnazione del Numero morfologico.

In quanto segue pertanto viene descritta la parte sperimentale della tesi, presentando in (2.1) la creazione della lista di nomi a dominanza plurale sulla base del corpus, in (2.2) la creazione del questionario e lo studio di *rating* e, infine, al paragrafo (2.3) sono illustrati i risultati ottenuti correlando lo studio di *rating* ai dati ricavati dal corpus.

2.1 Selezione dei nomi nel corpus ItWaC

Il lavoro sui corpora è stato effettuato sulla base dell'*Italian-Web-as-Corpus* (ItWaC), un corpus di testi in lingua italiana scaricati con metodi automatici dal web, accessibile dal sito <https://cqpweb.lancs.ac.uk/>.

ItWaC fa parte del progetto WaCky (*Web as Corpus kool ynitiative*), un insieme di corpora linguistici di cui fanno parte anche l'*ukWaC* per l'inglese e il *deWac* per il tedesco. I corpora contengono oltre il miliardo di parole e sono stati sviluppati tra il 2005 e il 2007 attraverso il *web crawling*, utilizzando cioè un programma per la raccolta e lo scaricamento di

pagine dal web.

Per costruire tali corpora, seguendo la metodologia esposta in Baroni et al. (2009), il primo passo consiste nell'identificare differenti URL che possano garantire una varietà sia per i contenuti, sia per il genere di documenti. Successivamente, in fase di *post-crawl*, le pagine selezionate vengono filtrate: in base al loro tipo (vengono mantenuti solo testi text/html); a seconda delle dimensioni (sono preservati solo documenti tra i 5 e i 200KB); e in base ai contenuti. Quest'ultimo processo prevede la rimozione dei codici (HTML e javascript) assieme a quelli che seguendo Fletcher (2004) vengono definiti *boilerplates*, ossia quelle parti di documento che tendono a ripetersi in modo simile attraverso le pagine e che si mostrano povere di contenuto testuale (come ad es. barre di navigazione, intestazioni, ecc.). In seguito, i testi ottenuti sono stati rapportati alle parole funzionali. In ciascuna di queste lingue, infatti, si suppone che vi sia un'alta proporzione tra parole funzionali e testo; pertanto le pagine in cui questa proporzione non era adeguata sono state scartate, considerandole non affidabili. Successivamente, anche le pagine pornografiche sono state eliminate in quanto contenevano lunghi testi ripetitivi, generati probabilmente per eludere i motori di ricerca. Inoltre, nei documenti sono state eliminate le porzioni in cui si incontravano numeri o materiale non linguistico, e in seguito anche le stesse parole funzionali sono state rimosse (411 in italiano). Infine, documenti in cui si sovrapponevano significative quantità di testo (*near-duplicates*) sono stati altresì scartati.

Relativamente all'ItWaC, questo lungo processo di filtraggio ha ridotto il materiale del corpus dagli 81 GB di partenza fino ai circa 10 GB finali di dati non compressi. Una volta compiuto questo lavoro, allo scopo di permettere e agevolare le future analisi di ricerca, è stato poi possibile etichettare il materiale del corpus con le differenti parti del discorso (attraverso il programma TreeTagger) e procedere alla lemmatizzazione dello stesso (utilizzando invece il programma Morph-it!). Bisogna tuttavia notare che il materiale così selezionato, pur parzialmente emendato, continua a contenere un'alta percentuale di errori, come la presenza di stringhe alfanumeriche o indirizzi e-mail.

Per quanto riguarda questa tesi, la prima fase della ricerca sull'ItWaC è stata volta a recuperare la frequenza di occorrenza di tutti i nomi presenti nel corpus, sia per la forma singolare che per quella plurale. Per fare ciò si sono utilizzate delle *query*, domande poste in sintassi CQP. Purtroppo non è stato possibile utilizzare una singola *query* per recuperare tutti i nomi del corpus (n=412.373.670) per motivi legati alla consultazione on-line. Pertanto le occorrenze dei nomi sono state raccolte spezzando le *query* per lettera alfabetica iniziale di

parola (al es. [word="ab.*" & pos="NOUN"]). Una volta ottenuti i risultati si è selezionata l'operazione *frequency breakdown* per ottenere i nomi ordinati per frequenza di occorrenza dal maggiore al minore. I nomi così ordinati sono poi stati copiati in Excel per facilitare le analisi. Un ulteriore problema è stato causato dall'impossibilità del server di ottenere i dati di *frequency breakdown* per quelle domande che ottengono come risultato nomi con oltre il milione di occorrenze, per un numero ristretto di nomi pertanto non è stato possibile ottenere la frequenza e sono quindi stati esclusi dallo studio.

In seguito, sono stati raggruppati i differenti nomi a seconda che avessero una, due, tre o quattro uscite per le forme singolari e plurali. Per semplificazione di analisi nomi che necessitavano al plurale della h diacritica per indicare il suono velare sono stati preliminarmente scartati (come ad es. amica/amiche, cieco/ciechi) ed alla stessa maniera non si sono utilizzati nomi terminanti al singolare con il digramma -io (del tipo crepaccio/crepacci). Al fine di conseguire dei risultati affidabili, solamente i termini a due o quattro uscite regolari con i suffissi trasparenti -a/-e oppure -o/-i sono stati presi in considerazione. Successivamente sono state eliminate le non-parole (cioè parole con refusi, in genere a bassissima frequenza, come ad es. zativo/zativi), i casi di omografia nome - verbo (ad es. regolo/regoli), nomi composti o derivati (asciugamano/asciugamani, riscaldamento/riscaldamenti). Inoltre si è tenuto conto della lunghezza ortografica, selezionando così termini che dessero una media di circa sei lettere per nome.

Dopo aver sommato la frequenza di occorrenza delle forme singolari e plurali, i nomi rimanenti sono stati divisi dividerli in: nomi ad alta frequenza di radice (> 20.000 occorrenze), a frequenza di radice media (< di 20.000 occorrenze e > di 10.000) e a bassa frequenza di radice (< di 10.000 occorrenze). Per ognuna di queste tre categorie si sono infine selezionati 96 nomi, suddivisi in nomi di tipo plural dominant femminili (con uscita -a/-e) (16) e maschili (uscita -o/-i) (16), singular dominant femminili (16) e maschili (16) e infine nomi a dominanza neutra suddivisi anch'essi in genere femminile (16) e maschile (16). In totale si sono ottenuti 576 nomi, presenti in appendice A.

2.2. Studio di rating

Una volta ottenuta la lista di nomi suddivisi per tipo di dominanza, è stato creato un questionario di frequenza soggettiva finalizzato a indagare la distribuzione dei singolari e plurali

nella lingua italiana secondo i parlanti nativi. In letteratura, si ritiene che la frequenza soggettiva sia una misura più affidabile della frequenza oggettiva, soprattutto se misurata in corpora di piccole dimensioni (i.e. Balota, Pilotti, & Cortese, 2001; Kuperman & Van Dyke, 2013; Williams & Morris, 2004).

Attraverso il sito <https://www.surveymonkey.com/> si sono realizzati due questionari da svolgere on-line intitolati “frequenza soggettiva di parole”. Ciascun questionario conteneva metà degli stimoli selezionati dal corpus, in modo tale che le forme singolari e plurali di ogni singolo nome fossero ripartiti fra i due questionari. Allo studio hanno partecipato 150 soggetti (femmine =94) distribuiti su tutto il territorio nazionale. L'età era compresa tra i 18 e i 69 anni (media =29 anni) e la scolarità era compresa tra gli 8 e i 21 anni. Requisiti essenziali per partecipare al questionario sono stati l'essere madrelingua italiana e aver ottenuto la maggiore età.

Ai soggetti è stato richiesto di rispondere secondo una scala di 7 valori riguardo alla frequenza con cui sentono/hanno sentito o leggono/hanno letto i nomi proposti. Le risposte possibili riguardo alle parole erano: 0 = mai letta o sentita; 1 = una o poche volte nella mia vita; 2 = una volta all'anno; 3 = una volta al mese; 4 = una volta alla settimana; 5 = una volta al giorno; 6 = più di una volta al giorno. Sedici parole a bassissima frequenza (i.e., meno di 215 occorrenze) sono state aggiunte a ciascun questionario come filler. I nomi nei questionari sono stati presentati ad ogni partecipante in un diverso ordine casuale.

2.3 Risultati

2.3.1 Distribuzione delle forme singolari e plurali

I risultati dello studio di rating per ogni singola occorrenza sono riportati nell'appendice A. I punteggi ottenuti riguardo alla frequenza soggettiva del singolare e il plurale per i diversi partecipanti allo studio sono schematizzati nella tabella 1 e illustrati negli istogrammi in figura 10, in confronto con la distribuzione degli stessi nomi nel corpus ItWaC.

Valori	Singolare	Plurale
$n = 0$	0	0
$0 < n < 1$	0	1
$1 < n < 2$	19	20
$2 < n < 3$	88	100
$3 < n < 4$	139	131
$4 < n < 5$	31	27
$n > 5$	7	5

Tabella 1. Riassunto dei risultati ottenuti nello studio di *rating*

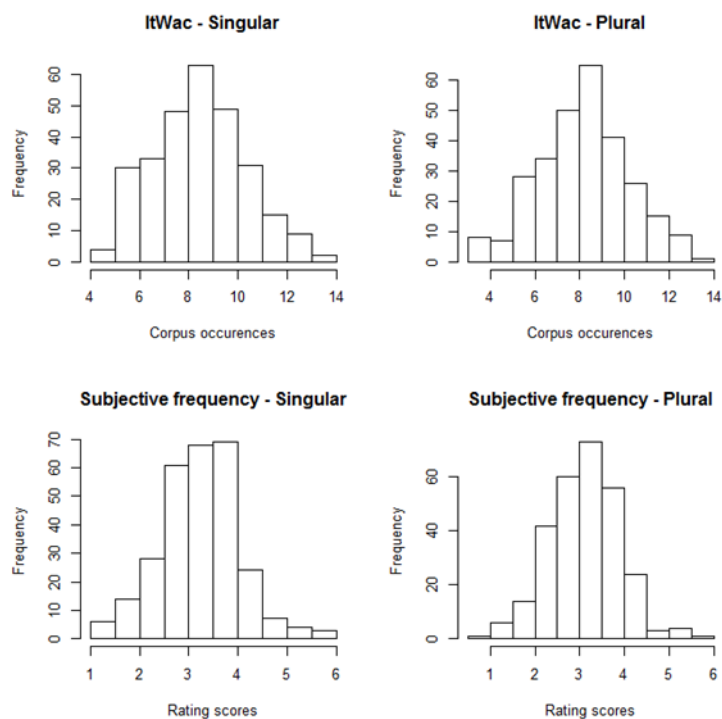


Figura 10. In alto: distribuzione dei nomi nel corpus ItWac. In basso: distribuzione dei risultati ottenuti nello studio di *rating*

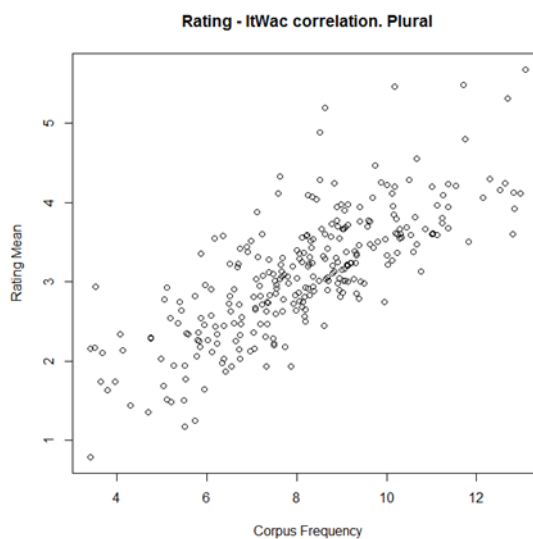
I soggetti non hanno assegnato valori pari a 0 agli stimoli somministrati, per cui ciascun nome presente nel questionario risulta essere udito o letto almeno una volta all'anno dai partecipanti. Come osservabile nei quattro istogrammi, la distribuzione dei nomi sia al singolare che al plurale è normale e non ci sono stati *bias* nella selezione dei nomi.

2.3.2 Correlazioni

Una prima analisi è stata quella di verificare se e quanto la frequenza soggettiva (ottenuta attraverso lo studio di *rating*) correla con la frequenza oggettiva (ottenuta attraverso il corpus ItWaC). I dati sono stati analizzati in R utilizzando il *test di Pearson*.

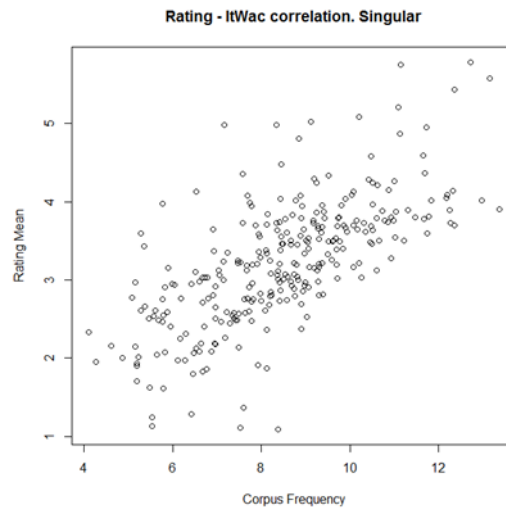
Dapprima si sono correlati i punteggi della frequenza soggettiva a quelli della frequenza oggettiva per quanto riguarda i nomi al plurale:

$$r(282) = 0.71, p < .001$$



In seguito si sono correlati i punteggi della frequenza soggettiva e quelli della frequenza oggettiva per quanto riguarda i nomi al singolare:

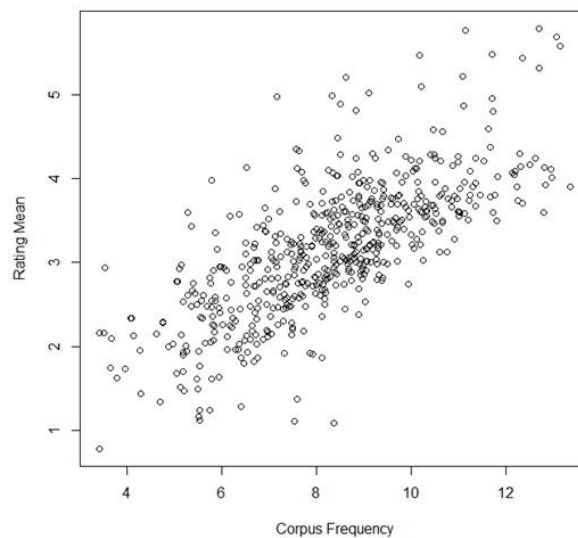
$$r(282) = 0.68, p < .001.$$



Per ultimo, si sono correlati i punteggi della frequenza soggettiva e quelli della frequenza oggettiva considerando tutte le occorrenze dei nomi (singolare e plurale):

$$r(566) = 0.72, p < .001$$

In tutti i casi la correlazione è risultata positiva e significativa, indipendentemente dalla misura di dominanza plurale.

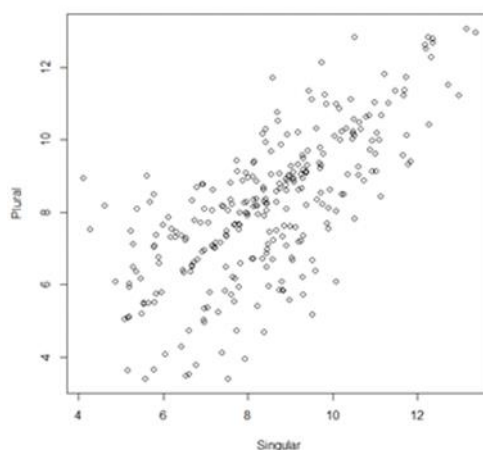


La seconda analisi invece era volta a verificare quanto l'occorrenza di un nome al singolare correli con l'occorrenza della corrispettiva forma al plurale. L'analisi è stata eseguita

sia rispetto alle frequenze nel corpus ItWaC sia rispetto ai punteggi ottenuti nello studio di *rating*. Anche in questo caso i dati sono stati analizzati in R utilizzando il *test di Pearson*.

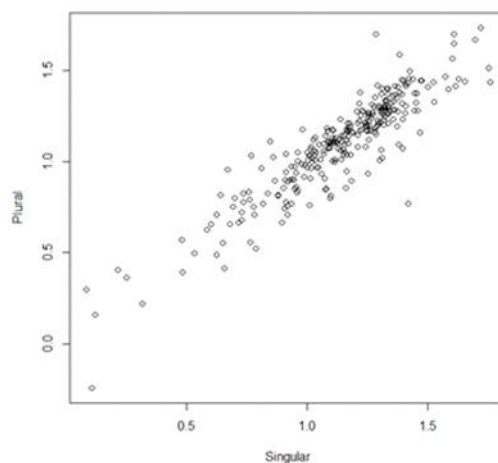
Per quanto riguarda i nomi nel corpus ItWaC la correlazione tra le forme singolari e plurali è stata:

$$r(282) = 0.70, p < .001$$



Mentre nello studio di *rating*:

$$r(282) = 0.91, p < .001$$



In entrambi i casi la correlazione è stata positiva e significativa. Nel caso del *rating*, tuttavia, questa correlazione è fortissima, descrivendo una differenza sostanzialmente nulla tra

la frequenza di occorrenza al singolare rispetto alla relativa forma plurale nel *rating*.

2.3.3. Misure di dominanza plurale

Un'ultima analisi effettuata riguarda in specifico il fenomeno della *plural dominance*. La misura di dominanza plurale è stata calcolata, seguendo la letteratura precedente (Baayen et al., 1996; 1997) come il rapporto frequenza al plurale/frequenza al singolare per quanto riguarda le occorrenze misurate nel corpus ItWaC, e come il rapporto tra frequenza soggettiva al plurale/frequenza soggettiva al singolare per quanto riguarda lo studio di *rating*. Tutti i rapporti sono presentati in appendice. In appendice sono stati evidenziati i 40 nomi più frequenti al singolare e i 40 più frequenti al plurale nel corpus ItWaC. Tuttavia, queste differenze di occorrenza singolare vs. plurale non sono state replicate nello studio di *rating* con il medesimo divario (cosa peraltro intuibile già osservando che la correlazione tra forme plurali e singolari nel *rating* è fortissima, come illustrato nel paragrafo precedente). In altre parole, nel *rating*, contrariamente a quanto osservato nel corpus, non è possibile stabilire una misura di dominanza plurale.

In parte ciò è riconducibile alla differenza di scala tra le frequenze del corpus, che non hanno un limite superiore diverso dal numero totale di item del corpus, e i punteggi del *rating*, che hanno invece un tetto limitato a $n = 6$. I valori restituiti dal *rating* sono pertanto più ravvicinati nel momento in cui viene calcolato il rapporto di occorrenza tra le due forme. Per questo motivo si è ritenuto più appropriato non confrontare direttamente le due misure nella correlazione.

Il primo scopo di questo studio è quello di fornire una lista di parole da poter utilizzare nello sviluppo di vari studi sperimentali avendo delle misure precise per la selezione degli item e l'interpretazione dei dati. Solo studi futuri potranno rivelare quale delle due misure sia più predittiva, anche rispetto al tipo di compito adottato.

Fatte queste premesse, va comunque notato che nel *rating* non è stato possibile osservare delle differenze anche minime tra le frequenze delle due forme. Questo fatto può essere imputato solo parzialmente alla scala ridotta di valori; ad esempio, in uno studio simile riguardante la morfologia di Numero è stato possibile rilevare delle differenze significative tra la frequenza soggettiva dei singolari e dei plurali (Zanini, Arcara e Franzon, 2014).

Capitolo 3. Discussioni e conclusioni

Riguardo ai risultati esposti nel capitolo 2 rispetto allo studio di *rating* e all'analisi del corpus ItWaC si possono sviluppare delle riflessioni di carattere metodologico. La forte correlazione tra singolare e plurale nello studio di *rating* e l'impossibilità di confrontare direttamente i dati provenienti dalle due misurazioni pone un problema di affidabilità: sono più predittive cioè le frequenze del corpus oppure quelle ottenute dallo studio di *rating*? Per rispondere a questa domanda è auspicabile eseguire ulteriori studi, ad esempio comportamentali, al fine verificare la forza predittiva dei due metodi di misurazione. E' altresì vero che studi precedenti non disponevano di tali misurazioni, e in merito ai dati discrepanti ottenuti in letteratura descritti nel capitolo 1, è lecito chiedersi se non siano frutto di scelte subottimali nella selezione degli items, dovute all'indisponibilità di corpora adeguati o al mancato utilizzo di tali misurazioni. In questo caso sarebbe opportuno replicare i vecchi paradigmi con nuovi stimoli e affrontare i risultati alla luce di paradigmi teorici più recenti. Risulta evidente che l'accuratezza metodologica permette di costruire esperimenti i cui dati risultino interpretabili, da qui l'importanza di prendere misure precise prima di qualsiasi studio.

Per quanto riguarda gli aspetti più teorici sia linguistici che psicolinguistici, non è stato possibile valutare se i nomi a dominanza plurale costituiscano o meno una classe, anche semantica, a parte. Infatti, non è possibile confrontare direttamente le misure di dominanza plurale ricavate dal corpus con quelle ricavate dallo studio di *rating* e quindi stabilire una misura coerente di dominanza plurale. Basandoci sulle indicazioni della frequenza soggettiva provenienti dallo studio di *rating*, pare che il fenomeno della dominanza non appartenga alla competenza dei parlanti. I parlanti infatti hanno mostrato scarsa sensibilità rispetto alla misura della dominanza (singolare o plurale che sia).

Questo dato può essere interpretato quindi a sfavore di modelli che postulano entrate separate a livello concettuale quanto del lemma nel caso dei nomi a dominanza plurale (vedi Levelt et al., 1999; Biedermann et al., 2012; Nickels et al., 2015) e a sfavore di analisi linguistiche che vedono nella flessione di plurale una sorta di derivazione (Booij, 1996). Se infatti la dominanza plurale fosse associata a un determinato tratto (come ad esempio "*multiple*" nei modelli di Levelt et al., 1999; e Nickels et al. 2015), esso dovrebbe appartenere alla competenza del parlante (ovvero alla sua grammatica universale) e avrebbe dovuto poter essere visibile nello studio *rating*. Invece, avendo riscontrato debole correlazione per la misura della *plural*

dominance, ne consegue che il fenomeno della dominanza plurale non può essere ritenuto secondo questi dati un fenomeno linguistico in senso stretto.

Il fatto di non dover postulare una doppia entrata semantica per i nomi *plural dominant* rende l'impostazione teorica più economica, non dovendo spiegare i meccanismi poco chiari secondo cui un determinato nome acquisirebbe un'ulteriore interpretazione semantica o entrata lemmatica. Nei lavori presentati nel primo capitolo infatti (Levelt et al., 1999; Biedermann et al. 2012; Nickels et al., 2015) non si fa menzione alla maniera secondo la quale nomi di tipo *plural dominant* riceverebbero un secondo lemma, ossia non viene indicata un'eventuale soglia di frequenza relativa o assoluta necessaria per cui una pluralizzazione a dominanza plurale sarebbe in grado di generare un'entrata propria. Allo stesso modo, in Levelt et al. (1999) non vengono forniti nemmeno dei criteri semantici particolari che distinguano i nomi a dominanza plurale dagli altri tipi di nomi. Questo fatto è di cruciale importanza, poiché se si ipotizzano per i nomi di tipo *plural dominant* delle entrate lessicali proprie significa sostanzialmente individuarli come una precisa categoria nominale indipendente nel lessico, piuttosto che come un insieme di nomi accumulati per una maggiore frequenza relativa nelle forme plurali.

Una visione di questo tipo, accosterebbe i nomi a dominanza plurale a dei plurali lessicali, forme in cui la pluralizzazione diventa parte inerente del lessema, cambiandone allo stesso tempo anche la struttura concettuale (i.e. Acquaviva, 2008). E' interessante notare però che le interpretazioni dei plurali lessicali possono essere ricondotte a un piccolo numero di denominatori semantici, quali ad esempio l'unità, l'identità, i confini o la forma, la coesione e l'esemplificazione. Queste nozioni semantiche riguardano nello specifico la maniera in cui un campo di referenze concettualizza le proprie parti costituenti. In particolare, si ipotizza che il significato di plurale non voglia dire "molti", ma "non-uno", e il determinato valore semantico di una pluralizzazione dipende da come un nome definisce e struttura il proprio "non-uno". In italiano ad esempio, "ossa", secondo l'ipotesi di Acquaviva (2002), rappresenterebbe una forma di plurale lessicale, differente dalla forma di tipo contabile "ossi" in quanto struttura le parti del proprio referente per il tratto [+coesione]. Per i nomi di tipo *plural dominant*, tuttavia, non è possibile riportare i diversi referenti a questo numero finito di valori semantici: le entità che designano, infatti, sfuggono (perlomeno in lingua italiana) da qualsiasi raggruppamento categoriale di tipo semantico. In più, a differenza dei plurali lessicali, il plurale nei nomi di tipo *plural dominant* non diventa parte del contenuto descrittivo di un nome e pertanto la pluralizzazione non determina un cambiamento nella struttura concettuale delle parti del referente. Contrariamente alle proposte di Levelt et al., (1999), dunque, appare poco

parsimonioso introdurre in un modello psicolinguistico un fenomeno ristretto come quello delle lessicalizzazioni plurali estendendolo all'intero insieme dei nomi di tipo *plural dominant*. Infatti, nonostante i concetti di plurale e singolare non siano la stessa identica cosa, essi sono comunque due concettualizzazioni molto vicine, quasi dei sinonimi, tanto che per la stragrande maggioranza dei nomi la pluralizzazione non determina un cambiamento della struttura concettuale del termine.

Facendo un parallelo con i valori di *mass* e *count*, si nota che nei nomi la contabilità può essere favorita da alcune caratteristiche del referente come ad esempio "essere un liquido", ma il valore di contabilità non può essere tuttavia predicibile a partire da proprietà ontologiche, lessicali e/o semantiche (i.e., Franzon et al, 2015; Zanini et al 2014). Allo stesso modo, anche il valore della dominanza plurale può essere favorito da alcune qualità del referente (come ad esempio il fatto di essere di piccole dimensioni), ma non è possibile associare per essi dei precisi tratti semantici. Date queste considerazioni, i risultati ottenuti nello studio di rating sembrano suggerire che il Numero morfologico venga assegnato non in base a tratti lessicali e/o semantici, ma contestualmente a seconda che vi sia o meno nel contesto comunicativo un'alta probabilità per un dato nome di occorre al plurale (non a caso molti dei nomi di tipo *plural dominant* sono dei duali).

I risultati di questo studio, invece, sono meglio interpretabili con alcune osservazioni presenti nello studio di Baayen et al. (2007). Nel lavoro appena citato infatti gli stimoli a dominanza singolare e plurale non sono stati trattati nell'analisi come due classi di nomi definiti, ma interpretati come stimoli che variano secondo una gradualità in base alla loro frequenza e divergenza rispetto al proprio paradigma flessivo. In questo schema pertanto non sono state supposte per i nomi di *tipo plural dominant* delle entrate semantiche diversificate, ipotizzando piuttosto che un solo approccio prettamente quantitativo potesse riflettere con più obiettività i dati ottenuti.

Per concludere, suggerendo che il fenomeno della *plural dominance* non sia un effetto presente nella grammatica del parlante, rimane da chiarire attraverso studi futuri se esso sia o non sia invece un fenomeno legato all'elaborazione linguistica. Più dettagliatamente, è conveniente appurare sia in comprensione che in produzione se le forme plurali dei nomi a dominanza plurale elicitino tempi di reazione più brevi rispetto alle forme singolari e plurali dei nomi *singular dominant* o se il fenomeno riscontrato in letteratura (per esempio in Baayen et al., 1996; 1997) non sia soltanto un effetto di frequenza in cui la interazione con la dominanza non gioca alcun ruolo. Dei dati in particolare presenti in questo studio riguardanti le correlazioni tra le

forme singolari e quelle plurali dei nomi nello studio di *rating* possono avvalorare quei modelli psicolinguistici (come ad esempio *il Parallel Dual-Route* o in parte il NDR) in cui la frequenza di radice gioca un ruolo centrale. Le correlazioni infatti indicano che tanto più un nome occorre al singolare tanto più esso occorre al plurale, anche laddove la frequenza di occorrenza della parola intera per la forma plurale non è fortemente attestata nel corpus. Lo stesso risultato si riscontra anche per quelle forme singolari poco attestate nel corpus che nello studio di *rating* sono state probabilmente avvantaggiate dall'alta frequenza delle corrispettive forme plurali. Questo risultato può suggerire pertanto che nei parlanti la frequenza di radice giochi un ruolo maggiore rispetto alla frequenza di occorrenza di parola, ma studi futuri nel campo dell'elaborazione lessicale con misure degli stimoli più accurate faranno luce su questo punto.

In ultima analisi, questo lavoro vuole mettere in evidenza come gli aspetti di elaborazione lessicale necessitino di essere presi in considerazione anche dalla linguistica formale, in quanto sono comunque parte della lingua e soprattutto perché un'interazione fra teoria e parte applicata può contribuire significativamente nell'impostare in maniera valida sia modelli predittivi formali che psicolinguistici. La collaborazione tra linguistica formale e linguistica sperimentale infatti è necessaria per comprendere al meglio ciò che pertiene alla lingua, ad ogni suo livello, distinguendo invece ciò che è il frutto di processi cognitivi generali che influiscono sugli atti linguistici. In questo studio si è cercato di far interagire queste due discipline con la speranza che da tale collaborazione possano essere individuati per il fenomeno della *plural dominance*, anche attraverso studi futuri, i punti di interfaccia che collegano fattori linguistici ad altri di ordine cognitivo.

Bibliografia

Acquaviva, P. (2002). Il plurale in -a come derivazione lessicale, in *Lingue e Linguaggio* 2002-2: 295-326.

Acquaviva, P. (2008). *Lexical Plurals: A Morphosemantic Approach: A Morphosemantic Approach*. OUP Oxford.

Aronoff, Mark. (1994). *Morphology by itself: Stems and inflectional classes*. No. 22. MIT press.

Baayen, H., Burani, C., & Schreuder, R. (1996). Effects of semantic markedness in the processing of regular nominal singulars and plurals in Italian. In *Yearbook of morphology 1996* (pp. 13-33). Springer Netherlands.

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language*, 37(1), 94-117.

Baayen, R., Levelt, W., Schreuder, R., & Ernestus, M. (2007, January). Paradigmatic structure in speech production. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society* (Vol. 43, No. 1, pp. 1-29). Chicago Linguistic Society.

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological review*, 118(3), 438.

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition*, 29(4), 639-647.

Baroni, Marco, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. "The WaCky wide web: a collection of very large linguistically processed web-crawled corpora." *Language resources and evaluation* 43, no. 3 (2009): 209-226.

Bender, M. Lionel. (1983). "Majang phonology and morphology". In *Nilo-Saharan Language Studies*, 114-147. East Lansing: Michigan State University.

Biedermann, B., Lorenz, A., Beyersmann, E., & Nickels, L. (2012). The influence of plural dominance in aphasic word production. *Aphasiology*, 26(8), 985-1004.

Biedermann, B., Beyersmann, E., Mason, C., & Nickels, L. (2013). Does plural dominance play a role in spoken picture naming? A comparison of unimpaired and impaired speakers. *Journal of Neurolinguistics*, 26(6), 712-736.

Booij, G. (1996). Inherent versus contextual inflection and the split morphology hypothesis. In *Yearbook of morphology 1995* (pp. 1-16). Springer Netherlands.

Butterworth, B. (1983) Lexical representation. In B. Butterworth (De.). *Language production* (Vol.2, pp.257-294). London: Academic Press.

Caramazza, A., Laudanna, A. & Romani, C. (1988) Lexical access and inflectional morphology. *Cognition*, 28, 297-332

Corbett, Greville G. (2000). *Number*. Cambridge Textbooks in Linguistics. Cambridge University Press.

Fletcher, W. (2004). Making the web more useful as a source for linguistic corpora. In U. Connor & T.Upton (Eds.), *Corpus linguistics in North America 2002* (pp. 191–205). Amsterdam: Rodopi

Franzon, F., Arcara, G., Riso, ML., & Zanini, C. (2015). Are mass nouns really different from count nouns? Some evidence from a lexical decision study. *AMLAP 21*, La Valletta, Malta, 3-5 settembre 2015.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: insights from connectionist models. *Psychological review*, 106(3), 491.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychological review*, 111(3), 662.

Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance*, 39(3), 802.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences*, 22(01), 1-38.

Luzzatti, C., Mondini, S. e Semenza, C. (2001). Lexical representation and processing of morphologically complex words: evidence from the reading performance of an Italian agrammatic patient. In: *Brain and Language* 79, 345-359.

Nickels, L., Biedermann, B., Fieder, N., & Schiller, N. O. (2015). The lexical-syntactic representation of number. *Language, Cognition and Neuroscience*, 30(3), 287-304.

Plaut, D. C., & Gonnerman, L. M. (2000). Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing?. *Language and Cognitive Processes*, 15(4-5), 445-485.

Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 131–154). Hillsdale, NJ: Erlbaum.

Seidenberg, M. S., & Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in cognitive sciences*, 4(9), 353-361.

Sereno, J. A., & Jongman, A. (1997). Processing of English inflectional morphology. *Memory & Cognition*, 25(4), 425-437.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication* (Urbana, IL).

Taft, M. & Forster, K.I. (1975) Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 14, 638-647.

Tiersma, P. M. (1982). Local and general markedness. *Language*, 832-849.

Zanini C., Arcara G., & Franzon F. (2014) Measuring the distribution of mass and count nouns. A comparison between a rating study and a corpus based analysis, *Palc 14*, Łódź, Poland DOI: 10.13140/RG.2.1.3986.5448

Williams, R. S., & Morris, R. K. (2004). Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology*, 16(1-2), 312-339.

Appendice A

Item_Sg	Corpus_Frequency	Media_Rating	Item_Pl	Corpus_Frequency	Media_Rating	Plural dominance_Corpus	Plural dominance_Rating
acaro	318	2.7538	acari	1142	2.8125	3.5912	1.0213
accisa	451	1.9714	accise	2597	1.9265	5.7583	0.9772
aceto	681	4.1296	aceti	33	2.1613	0.0485	0.5234
acqua	333623	5.7879	acque	102627	4.2105	0.3076	0.7275
afasia	608	1.2857	afasie	73	1.4394	0.1201	1.1195
afta	740	2.0882	afte	115	2.2836	0.1554	1.0935
agenzia	51607	3.7368	agenzie	43260	3.4727	0.8383	0.9293
aiuola	1147	3.0597	aiuole	2234	3.0923	1.9477	1.0107
albero	40994	4.2128	alberi	36216	4.2857	0.8834	1.0173
allergia	2857	3.5593	allergie	4028	3.5179	1.4099	0.9884
allodola	533	1.9714	allodole	1516	1.9275	2.8443	0.9777
amido	1059	2.9231	amidi	213	2.4769	0.2011	0.8474
amuleto	811	2.4058	amuleti	759	2.2537	0.9359	0.9368
anatra	2121	3.1231	anatre	2125	3.1343	1.0019	1.0036
ancella	796	1.8261	ancelle	691	1.9286	0.8681	1.0561
aneddoto	2718	3.2029	aneddoti	3885	3.2273	1.4294	1.0076
anelito	1988	1.3714	aneliti	311	1.2464	0.1564	0.9088
angelo	17499	3.4918	angeli	15243	3.4762	0.8711	0.9955
anomalia	10877	3.5246	anomalie	9790	3.2381	0.9001	0.9187
antenna	8583	3.5323	antenne	9129	3.2167	1.0636	0.9107
aquila	5733	2.9848	aquile	1835	3.0149	0.3201	1.0101
aratro	1205	2.4627	aratri	332	2.3571	0.2755	0.9571
arazzo	678	2.1324	arazzi	1483	2.3043	2.1873	1.0807
asilo	55415	3.2759	asili	9282	3.2063	0.1675	0.9788
asintoto	239	1.6212	asintoti	182	1.4783	0.7615	0.9118
assenza	136966	4.0196	assenze	12149	3.9483	0.0887	0.9823
astro	3635	2.7941	astri	4382	2.8824	1.2055	1.0316
ateneo	12040	2.8154	atenei	11138	2.8551	0.9251	1.0141
attrezzo	2796	3.5873	attrezzi	8839	3.6852	3.1613	1.0273
aula	59851	4.2593	aule	15077	3.7593	0.2519	0.8826
avaria	2070	2.5821	avarie	504	2.3429	0.2435	0.9073
azienda	222833	4.1429	aziende	218145	4.2963	0.9790	1.0370
baia	5309	2.7463	baie	819	2.1429	0.1543	0.7803
barba	13660	4.3396	barbe	794	3.1833	0.0581	0.7336
barriera	14878	3.4923	barriere	30578	3.5556	2.0552	1.0181
bavero	862	1.8657	baveri	44	1.6286	0.0510	0.8729
belva	1740	3.2154	belve	1593	3.1212	0.9155	0.9707

bestia	11540	3.6032	bestie	9163	3.2034	0.7940	0.8890
bevanda	6420	4.0196	bevande	19537	4.2593	3.0431	1.0596
biglia	353	2.5942	biglie	867	2.5507	2.4561	0.9832
biglietto	35590	4.5778	biglietti	26304	4.2037	0.7391	0.9183
biscotto	1294	4.9778	biscotti	5576	5.2000	4.3091	1.0446
bomba	22902	4.0862	bombe	24644	4.1200	1.0761	1.0083
bordo	60733	3.5472	bordi	9244	3.7222	0.1522	1.0493
calamaro	2424	2.9559	calamari	3000	3.0435	1.2376	1.0296
camelia	178	1.7059	camelie	380	1.6429	2.1348	0.9631
camera	69057	5.7561	camere	43273	4.5556	0.6266	0.7914
cancro	23629	3.6949	cancri	442	2.9016	0.0187	0.7853
candela	7160	3.5606	candele	6386	3.5690	0.8919	1.0023
canguro	891	2.7681	canguri	823	2.7164	0.9237	0.9813
canoa	3236	2.6143	canoe	836	2.3284	0.2583	0.8906
carciofo	1257	3.2353	carciofi	3152	3.3582	2.5076	1.0380
cardo	648	2.0714	cardi	569	1.9714	0.8781	0.9517
carogna	786	2.7164	carogne	677	2.6308	0.8613	0.9685
carota	3382	3.7083	carote	4215	4.0690	1.2463	1.0972
cellula	13404	3.3333	cellule	68045	3.5873	5.0765	1.0762
chiazza	665	3.1045	chiazze	1393	3.0746	2.0947	0.9904
chiodo	4347	3.1167	chiodi	4101	3.4286	0.9434	1.1001
chitarra	31069	3.6129	chitarre	8587	3.1667	0.2764	0.8765
cielo	114109	4.5952	cieli	14525	3.7742	0.1273	0.8213
cilindro	6275	2.7910	cilindri	5855	3.0597	0.9331	1.0963
cimitero	14976	3.5000	cimiteri	4272	3.0606	0.2853	0.8745
collana	17267	3.8000	collane	3741	3.5938	0.2167	0.9457
corallo	2172	2.7647	coralli	2148	2.6232	0.9890	0.9488
corazza	2455	2.7500	corazze	732	2.4769	0.2982	0.9007
corsia	10253	3.8596	corsie	8610	3.9000	0.8398	1.0105
cosa	514140	5.5758	cose	480646	5.6833	0.9349	1.0193
cravatta	4752	3.4655	cravatte	1244	3.3077	0.2618	0.9545
crystallo	8703	3.1613	cristalli	6896	3.1270	0.7924	0.9891
crosta	4558	3.3793	croste	668	3.2308	0.1466	0.9560
delitto	25907	3.6441	delitti	22018	3.3333	0.8499	0.9147
detrito	292	2.4848	detriti	3960	2.8333	13.5616	1.1402
diacono	1798	2.5714	diaconi	1554	2.4638	0.8643	0.9581
dialetto	11645	3.9583	dialetti	3652	3.5846	0.3136	0.9056
difetto	36934	4.2400	difetti	25302	3.9583	0.6851	0.9336
dinastia	7292	2.6923	dinastie	1185	2.6714	0.1625	0.9922
diritto	637364	3.9000	diritti	427987	4.1111	0.6715	1.0541
disamina	4327	1.0882	disamine	110	1.3478	0.0254	1.2385

divano	9074	5.0208	divani	1985	4.1186	0.2188	0.8203
dogana	3776	2.8088	dogane	2865	2.8254	0.7587	1.0059
dollaro	12434	3.7925	dollari	85924	3.9474	6.9104	1.0408
donna	233953	5.4364	donne	327764	5.3103	1.4010	0.9768
dosso	4627	3.5593	dossi	1081	3.5156	0.2336	0.9877
effetto	194154	4.0435	effetti	307410	4.2453	1.5833	1.0499
eresia	4903	3.0154	eresie	1764	3.1029	0.3598	1.0290
esodo	8308	2.5303	esodi	857	2.4627	0.1032	0.9733
etnia	6264	3.2154	etnie	6583	3.2090	1.0509	0.9980
ettaro	4497	2.7286	ettari	20921	2.7424	4.6522	1.0051
fagiolo	1020	3.3676	fagioli	6516	3.7344	6.3882	1.1089
fanfara	638	1.7971	fanfare	608	1.8714	0.9530	1.0414
fantasia	45231	3.8824	fantasie	7294	3.7679	0.1613	0.9705
farfalla	5806	3.4559	farfalle	3956	3.3077	0.6814	0.9571
farmaco	23903	4.1250	farmaci	60062	4.1964	2.5127	1.0173
fegato	13573	3.4915	fegati	178	2.5397	0.0131	0.7274
fessura	2415	3.1940	fessure	2115	3.0615	0.8758	0.9585
fiaba	7831	3.3433	fiabe	8088	3.1475	1.0328	0.9415
fiamma	12256	3.8727	fiamme	16372	3.4262	1.3358	0.8847
fiato	19050	3.7000	fiati	2217	2.7761	0.1164	0.7503
finanza	58717	3.7586	finanze	62024	3.6000	1.0563	0.9578
flotta	10865	2.8030	flotte	1573	2.7391	0.1448	0.9772
flusso	36062	3.4603	flussi	39529	3.3729	1.0961	0.9747
fobia	1391	3.3492	fobie	1302	2.9524	0.9360	0.8815
fragola	1948	3.7258	fragole	4256	3.5397	2.1848	0.9500
fronda	960	2.0882	fronde	2272	2.1818	2.3667	1.0448
fronzolo	71	1.9565	fronzoli	1853	2.5970	26.0986	1.3274
frottola	197	2.6143	frottole	660	2.7273	3.3503	1.0432
galassia	8568	2.9077	galassie	7638	2.8906	0.8915	0.9941
galera	9699	3.1552	galere	1324	2.8209	0.1365	0.8941
gambero	1842	3.2239	gamberi	3175	3.2923	1.7237	1.0212
gemma	2027	2.7500	gemme	3785	2.9118	1.8673	1.0588
gioiello	8108	3.5000	gioielli	10843	3.3898	1.3373	0.9685
giungla	6224	2.8841	giungle	353	2.5441	0.0567	0.8821
globulo	214	2.6667	globuli	3327	3.2500	15.5467	1.2188
glossa	254	1.2429	glosse	242	1.5000	0.9528	1.2069
gola	19841	4.0377	gole	1886	2.9194	0.0951	0.7230
golfo	6861	3.0000	golfi	342	2.2500	0.0498	0.7500
graffito	486	2.7612	graffiti	1887	2.9538	3.8827	1.0698
grafia	2124	3.1875	grafie	256	2.3529	0.1205	0.7382
grillo	5356	3.0448	grilli	1373	2.7121	0.2563	0.8907

grissino	210	3.4308	grissini	577	3.5763	2.7476	1.0424
guanto	2607	3.6964	guanti	7278	3.9455	2.7917	1.0674
impresa	207105	3.8983	imprese	375846	3.9245	1.8148	1.0067
incognita	2924	3.2879	incognite	2715	3.1935	0.9285	0.9713
incubo	16526	3.6897	incubi	5148	3.6607	0.3115	0.9922
indulto	2774	1.9130	indulti	52	1.7391	0.0187	0.9091
inerzia	10685	2.9846	inerzie	502	2.2239	0.0470	0.7451
infarto	8399	3.4848	infarti	806	3.2239	0.0960	0.9251
insetto	3448	3.8421	insetti	12161	3.7581	3.5270	0.9781
intestino	4836	3.4590	intestinali	386	2.9545	0.0798	0.8542
labirinto	6659	3.0441	labirinti	1530	2.7500	0.2298	0.9034
lacrima	4449	3.7385	lacrime	29676	3.6607	6.6703	0.9792
lacuna	5096	3.0758	lacune	8467	3.0152	1.6615	0.9803
lampada	7154	4.0800	lampade	7463	3.6981	1.0432	0.9064
larva	1048	2.6567	larve	3331	2.6812	3.1784	1.0092
lastra	4276	3.0597	lastre	5658	3.0299	1.3232	0.9902
lessico	6607	3.1970	lessici	446	2.5758	0.0675	0.8057
lingotto	236	2.5072	lingotti	485	2.4328	2.0551	0.9703
locusta	130	2.0000	locuste	447	2.1176	3.4385	1.0588
lombo	179	1.9000	lombi	412	2.2647	2.3017	1.1920
luna	36529	3.9615	lune	2492	2.9851	0.0682	0.7535
lustro	1779	2.1429	lustri	1789	2.2029	1.0056	1.0280
maceria	60	2.3333	macerie	7655	3.0462	127.5833	1.3055
malattia	118272	4.3673	malattie	87220	4.2373	0.7375	0.9702
marito	67282	4.8696	mariti	4652	4.0377	0.0691	0.8292
massaia	1056	2.1765	massaie	1062	2.1194	1.0057	0.9738
materasso	2377	3.9464	materassi	2151	3.3051	0.9049	0.8375
matrigna	1594	2.5072	matrigne	62	2.1343	0.0389	0.8513
melodia	8585	3.4032	melodie	7336	3.2462	0.8545	0.9538
mensola	1002	3.6491	mensole	1005	3.3793	1.0030	0.9261
mercato	433345	4.0167	mercati	75781	3.7368	0.1749	0.9303
metallo	15856	3.5333	metalli	11769	3.4603	0.7422	0.9793
metodo	117677	3.7800	metodi	76159	3.8000	0.6472	1.0053
miliardo	16947	3.7931	miliardi	186349	4.0638	10.9960	1.0714
miniera	7541	2.8551	miniere	7438	2.9851	0.9863	1.0455
minuzia	335	2.0725	minuzie	318	2.0580	0.9493	0.9930
mirto	1042	2.1940	mirti	154	1.6857	0.1478	0.7683
moccioso	159	2.7727	mocciosi	156	2.7727	0.9811	1.0000
monumento	16125	3.3906	monumenti	18037	3.5077	1.1186	1.0345
morbo	7836	2.9242	morbi	265	2.3382	0.0338	0.7996
muscolo	5089	3.7800	muscoli	16083	4.0612	3.1603	1.0744

museo	42492	3.5000	musei	29643	3.5455	0.6976	1.0130
mutanda	322	3.9800	mutande	4981	4.8913	15.4689	1.2290
narciso	384	2.4058	narcisi	329	2.2647	0.8568	0.9414
nausea	6661	3.4483	nausee	352	3.3548	0.0528	0.9729
nembo	254	1.1286	nembi	246	1.1714	0.9685	1.0380
nervo	3311	3.3833	nervi	8014	3.9800	2.4204	1.1764
nodo	30183	3.7288	nodi	30279	3.6032	1.0032	0.9663
norma	231659	3.7000	norme	364481	3.5968	1.5734	0.9721
notizia	122865	4.9565	notizie	125423	4.7955	1.0208	0.9675
ombra	55140	3.7966	ombre	21473	3.5357	0.3894	0.9313
omelia	3358	1.8676	omelie	840	2.0286	0.2501	1.0862
ordigno	4466	2.9697	ordigni	4302	2.9385	0.9633	0.9895
organo	73996	3.5082	organi	135391	3.5000	1.8297	0.9977
orma	1002	2.8095	orme	6564	3.1077	6.5509	1.1061
otorino	262	2.5429	otorini	30	2.1594	0.1145	0.8492
pannello	10939	3.1667	pannelli	13429	2.9844	1.2276	0.9424
pantofola	194	3.5965	pantofole	1236	3.8750	6.3711	1.0774
papaia	322	2.4706	papaie	39	2.1000	0.1211	0.8500
paradiso	19842	3.6667	paradisi	3392	2.7419	0.1710	0.7478
parcella	1362	2.5942	parcelle	1129	2.3636	0.8289	0.9111
petalo	793	3.0299	petali	4193	3.2353	5.2875	1.0678
pigna	172	2.9683	pigne	165	2.9242	0.9593	0.9852
pilastro	6220	3.0735	pilastrri	8705	3.0149	1.3995	0.9809
piscina	20689	3.6182	piscine	5592	3.2951	0.2703	0.9107
poligono	3533	2.6818	poligoni	3575	2.5000	1.0119	0.9322
pollo	9786	4.2963	polli	5598	3.6032	0.5720	0.8387
polso	7355	3.7895	polsi	3331	3.5593	0.4529	0.9393
portico	3335	3.3443	portici	3498	3.3651	1.0489	1.0062
prato	11626	3.6780	prati	8865	3.3788	0.7625	0.9187
prestito	34557	3.4848	prestiti	22605	3.2105	0.6541	0.9213
quaderno	4648	4.4783	quaderni	3828	4.0962	0.8236	0.9147
quadro	212941	3.7333	quadri	34162	3.6842	0.1604	0.9868
quarzo	1470	2.4493	quarzi	191	1.9420	0.1299	0.7929
quesito	26856	3.2131	quesiti	27671	3.3651	1.0303	1.0473
ramarro	172	2.1449	ramarri	38	1.7429	0.2209	0.8125
rampa	4492	3.2031	rampe	3803	3.1912	0.8466	0.9963
rampollo	759	2.1857	rampolli	586	2.0290	0.7721	0.9283
rata	10747	3.6610	rate	14391	3.6949	1.3391	1.0093
reliquia	1663	2.4853	reliquie	3534	2.5652	2.1251	1.0322
reperto	2256	2.9063	reperti	12437	3.0000	5.5129	1.0323
risorsa	36919	3.6379	risorse	373648	4.1296	10.1208	1.1352

riviera	3704	2.8485	riviere	227	2.6364	0.0613	0.9255
rovo	474	2.2500	rovi	1510	2.6232	3.1857	1.1659
ruspa	616	2.9516	ruspe	1588	3.2787	2.5779	1.1108
sabbia	23586	3.1667	sabbie	3090	2.8615	0.1310	0.9036
sandalo	839	3.0294	sandali	2426	3.0635	2.8915	1.0112
sasso	7691	3.6500	sassi	8378	3.6667	1.0893	1.0046
sbornia	1143	3.1212	sbornie	218	2.7500	0.1907	0.8811
scaglia	394	2.9531	scaglie	2101	3.2188	5.3325	1.0899
scarpa	4142	4.9811	scarpe	26408	5.4694	6.3757	1.0980
scettro	2244	2.6000	scettri	115	2.2899	0.0512	0.8807
scimmia	4278	3.6333	scimmie	5935	3.4167	1.3873	0.9404
scogliera	2911	2.8235	scogliere	2935	2.6364	1.0082	0.9337
scoria	270	2.6119	scorie	8260	2.8615	30.5926	1.0956
scrigno	2379	2.4776	scrigni	379	2.4559	0.1593	0.9912
scrivania	10405	4.2407	scrivanie	1363	3.6032	0.1310	0.8497
scrupolo	3626	3.2462	scrupoli	7099	3.2985	1.9578	1.0161
scudo	10601	3.1194	scudi	5874	2.8939	0.5541	0.9277
sfarzo	1047	2.5072	sfarzi	144	2.0299	0.1375	0.8096
sigaro	2371	2.7206	sigari	1878	2.8551	0.7921	1.0494
simbolo	61611	3.8750	simboli	26975	3.6102	0.4378	0.9317
sindaco	124939	3.5968	sindaci	24933	3.2742	0.1996	0.9103
sintomo	9518	3.7586	sintomi	27304	3.7966	2.8687	1.0101
slavina	282	2.0435	slavine	246	1.9429	0.8723	0.9508
soglia	40328	3.1231	soglie	10550	3.0328	0.2616	0.9711
soldato	18884	3.4000	soldati	59860	3.6034	3.1699	1.0598
soldo	5316	3.6111	soldi	122470	5.4808	23.0380	1.5178
sopruso	1600	2.5821	soprusi	3589	2.6462	2.2431	1.0248
sorso	2164	4.0784	sorsi	478	3.5484	0.2209	0.8700
sostanza	95551	3.8036	sostanze	85734	3.6724	0.8973	0.9655
spada	28405	3.0303	spade	4920	3.0152	0.1732	0.9950
spalla	18072	3.9600	spalle	76828	4.0926	4.2512	1.0335
spalto	101	2.1571	spalti	3559	2.8143	35.2376	1.3046
spatola	733	2.9846	spatole	34	2.9385	0.0464	0.9845
spirito	129848	3.8070	spiriti	11145	3.3529	0.0858	0.8807
sponda	11637	3.2154	sponde	9347	3.0000	0.8032	0.9330
spora	187	2.0143	spore	1804	2.2239	9.6471	1.1041
sprazzo	535	2.3077	sprazzi	1766	2.2857	3.3009	0.9905
spugna	4586	4.0370	spugne	972	3.4375	0.2119	0.8515
squalo	3645	3.0870	squali	3585	2.9286	0.9835	0.9487
stanza	65239	5.2128	stanze	22314	4.2174	0.3420	0.8091
statua	16255	3.2000	statue	10026	3.2424	0.6168	1.0133

stella	33242	4.2830	stelle	67766	3.9615	2.0386	0.9249
steppa	1317	2.2609	steppe	1153	2.1571	0.8755	0.9541
sterlina	873	3.0317	sterline	6014	3.0159	6.8889	0.9948
strategia	80936	3.8983	strategie	61332	3.6102	0.7578	0.9261
strumento	197975	4.0847	strumenti	274686	4.1633	1.3875	1.0192
tariffa	25571	3.7586	tariffe	52388	3.6667	2.0487	0.9755
tartufo	4301	3.0149	tartufi	2753	2.7246	0.6401	0.9037
taverna	1761	3.2500	taverne	666	2.8235	0.3782	0.8688
tazza	6921	4.8095	tazze	2059	4.3333	0.2975	0.9010
tifoso	6035	3.8333	tifosi	37683	3.5862	6.2441	0.9355
timpano	1260	3.0000	timpani	1199	3.0308	0.9516	1.0103
tono	35247	3.6852	toni	27882	3.5968	0.7910	0.9760
torsolo	417	2.9412	torsoli	59	2.3429	0.1415	0.7966
tragedia	40462	3.7903	tragedie	8357	3.6667	0.2065	0.9674
traghetto	4570	2.8676	traghetti	3174	2.7536	0.6945	0.9602
transenna	337	2.9077	transenne	1619	2.8308	4.8042	0.9735
tregua	10982	3.2031	tregue	309	2.8209	0.0281	0.8807
trono	14662	3.2576	troni	590	2.4478	0.0402	0.7514
truppa	5857	2.9242	truppe	46953	3.1270	8.0166	1.0693
turno	52991	4.1538	turni	16922	4.4667	0.3193	1.0753
uccello	7414	3.9464	uccelli	25880	3.8519	3.4907	0.9760
ugello	179	1.9286	ugelli	166	1.5143	0.9274	0.7852
ulcera	1536	2.5522	ulcere	1313	2.4697	0.8548	0.9677
unghia	1951	4.3519	unghie	6834	4.2407	3.5028	0.9745
urna	3363	2.3636	urne	11873	2.7857	3.5305	1.1786
vacanza	47892	3.7593	vacanze	41738	3.8136	0.8715	1.0144
vaso	9653	3.7321	vasi	11247	3.3333	1.1651	0.8931
velivolo	7343	2.3788	velivoli	5499	2.4493	0.7489	1.0296
vena	10693	3.6935	vene	7709	3.5500	0.7209	0.9611
vespro	323	1.6176	vespri	251	1.7714	0.7771	1.0951
vettura	16486	3.0167	vetture	10973	3.2381	0.6656	1.0734
viatico	1851	1.1143	viatici	30	0.7826	0.0162	0.7023
vicolo	5789	3.5079	vicoli	5410	3.3279	0.9345	0.9487
vigneto	2966	2.7313	vigneti	7878	2.8116	2.6561	1.0294
violino	8080	2.9688	violini	2041	3.0299	0.2526	1.0206
vocabolo	2824	3.3651	vocaboli	2992	3.4000	1.0595	1.0104
zampa	2277	3.9815	zampe	9282	3.9808	4.0764	0.9998
zanna	324	2.5441	zanne	1186	2.6471	3.6605	1.0405
zavorra	1726	2.4853	zavorre	345	2.1791	0.1999	0.8768
zigomo	364	3.1538	zigomi	741	2.9048	2.0357	0.9210
zucchero	27171	5.0889	zuccheri	4941	4.2889	0.1818	0.8428

zuppa	4147	3.7302	zuppe	840	3.4167	0.2026	0.9160
-------	------	--------	-------	-----	--------	--------	--------

I 40 nomi con la maggiore plural dominance

	Plural dominance_Corpus
macerie	127.5833
spalti	35.2376
scorie	30.5926
fronzoli	26.0986
soldi	23.0380
globuli	15.5467
mutande	15.4689
detriti	13.5616
miliardi	10.9960
risorse	10.1208
spore	9.6471
truppe	8.0166
dollari	6.9104
sterline	6.8889
lacrime	6.6703
orme	6.5509
fagioli	6.3882
scarpe	6.3757
pantofole	6.3711
tifosi	6.2441
accise	5.7583
reperti	5.5129
scaglie	5.3325
petali	5.2875
cellule	5.0765
transenne	4.8042
ettari	4.6522
biscotti	4.3091
spalle	4.2512
zampe	4.0764
graffiti	3.8827
zanne	3.6605

i 40 nomi con la minor plural dominance

	Plural dominance_Corpus
torsolo	0.1415
sfarzo	0.1375
galera	0.1365
sabbia	0.1310
scrivania	0.1310
quarzo	0.1299
cielo	0.1273
papaia	0.1211
grafia	0.1205
afasia	0.1201
fiato	0.1164
otorino	0.1145
esodo	0.1032
infarto	0.0960
gola	0.0951
assenza	0.0887
spirito	0.0858
intestino	0.0798
marito	0.0691
luna	0.0682
lessico	0.0675
riviera	0.0613
barba	0.0581
giungla	0.0567
nausea	0.0528
scettro	0.0512
bavero	0.0510
golfo	0.0498
aceto	0.0485
inerzia	0.0470
spatola	0.0464
trono	0.0402

acari	3.5912	matrigna	0.0389
urne	3.5305	morbo	0.0338
insetti	3.5270	tregua	0.0281
unghie	3.5028	disamina	0.0254
uccelli	3.4907	indulto	0.0187
locuste	3.4385	cancro	0.0187
frottole	3.3503	viatico	0.0162
sprazzi	3.3009	fegato	0.0131