

LA FORMAZIONE DI PAROLE IN UN CORPUS DIACRONICO: APPUNTI SULLA PRODUTTIVITÀ DEI SUFFISSI *-MENTO* E *-ZIONE* NEL CINQUECENTO

PAVEL ŠTICHAUER

Università Carlo di Praga

WORD-FORMATION IN A DIACHRONIC CORPUS: REMARKS ON THE PRODUCTIVITY OF THE SUFFIXES *-MENTO* AND *-ZIONE* IN THE ITALIAN LANGUAGE OF THE 16TH CENTURY

This paper deals with two deverbal suffixes, *-mento* and *-zione*, in the Italian language of *Cinquecento*. The article aims to show, within the corpus-based quantitative approach to morphological productivity, that in the 16th century the suffix *-zione* cannot be considered a fully *available* means of word-formation but should be rather seen as a result of a massive *borrowing* process. The data, drawn from a 10 million corpus of *Cinquecento* (being part of the large LIZ 4.0 corpus), are presented along with the discussion of the type frequency, token frequency, vocabulary growth and the *P* values. The high token frequency of *-zione* is highlighted with special respect to the present-day situation where the suffix is the “frequency topscorer”. However, the main difference between the 16th century and the present-day *-zione* formations is stressed, i.e. the absence of nouns with *-izzazione* (and the scarce presence of *-ificazione*) which is due to the absence of causative verbs in *-izzare* and *-ificare*. It is only later, as witnessed by three major dictionaries, that these causative suffixes settle down as really *available* becoming thus the principal *input* to the suffix *-zione*.

Key words: diachronic corpus – word-formation – morphological productivity – italian language of the 16th century – suffixes *-mento* and *-zione*

Parole chiave: corpus diacronico – formazione delle parole – produttività morfologica – italiano del *Cinquecento* – suffissi *-mento* e *-zione*

1. Introduzione

In questo articolo mi propongo di prendere nuovamente in esame la situazione cinquecentesca dei due suffissi deverbali *-mento* e *-zione* che avevo già in parte trattato in un lavoro precedente (cfr. Štichauer 2009).¹ Ho intenzione di analizzare il quadro dei

¹ In alcuni passi riprendo, seppur con qualche modifica, il testo già pubblicato come Štichauer 2009 (mi baso, in particolare, sul capitolo 4.6, pp. 116–128). Il presente saggio rientra nell’ambito del progetto di ricerca GAČR P406/12/0450 (*La formazione delle parole in italiano dal Cinquecento ai giorni nostri*), finanziato dalla *Czech Science Foundation*; ed è ugualmente *output* del progetto “Program rozvoje

due suffissi, considerandolo non come un punto di arrivo (posizione propria del lavoro citato), ma piuttosto come il punto di partenza di un processo che si sarebbe sviluppato e assestato appieno soltanto nei periodi successivi. Cerco infatti di mostrare come la presunta grande produttività del suffisso *-zione* sia piuttosto, ancora in questo periodo, un vasto processo di *borrowing* morfologico (cfr. Matras 2009: 209–212) anziché un procedimento di formazione di parole autonomo e pienamente *disponibile* (nel senso di Corbin 1987).

L'articolo è organizzato nel modo seguente: in (2) introduco brevemente l'aspetto quantitativo e qualitativo dei due suffissi (e della formazione di parole in generale); in (3) presento i dati elaborati a partire dal corpus LIZ 4.0; in (4) mi soffermo sul gruppo di formazioni più importante: quello degli *hapax legomena*; in (5), invece, presento l'aspetto quantitativo della produttività dei due suffissi; in (6) cerco di descrivere la struttura morfologica del suffisso *-zione* nel Cinquecento, mentre in (7) riassumo i risultati presentati.

2. Aspetti qualitativi e quantitativi

Dal punto di vista dei processi di formazione di parole, le questioni più importanti sono generalmente quelle che riguardano l'aspetto qualitativo e quantitativo di un dato processo morfologico.

Per l'aspetto qualitativo si può intendere il numero e la natura delle *restrizioni* che regolano l'applicabilità alle basi che ne costituiscono l'*input* principale. Per l'aspetto quantitativo, invece, si deve intendere l'estensione o anche una specie di "copertura" della categoria morfologica (o della regola di formazione di parole, d'ora in poi RFP) dei mezzi in questione. Tale estensione può essere concepita sia dal punto di vista del numero totale dei *lemmi* a cui la RFP ha dato luogo, sia dal punto di vista della *frequenza delle occorrenze* dei lemmi, o anche dal punto di vista del *potenziale accrescimento* che di solito coincide con la nozione di produttività.

2.1 L'aspetto qualitativo

Se si prende come punto di riferimento il quadro delle restrizioni vigenti nell'italiano contemporaneo, si può sostenere che i suffissi *-mento* e *-zione* siano, in generale, in distribuzione complementare.

Le restrizioni, che ormai sono ben note (cfr. Thornton 1987–1988: 310–358; Scalise 1994: 106, 114–115; Gaeta 2004: 327–334), sono di tipo fonologico e di tipo morfologico che riguardano la classe flessiva (coniugazione), la base semplice versus la base complessa, la presenza, nella base complessa, di suffissi o prefissi che condizionano la selezione di uno o dell'altro dei due suffissi.

vědních oblastí na Univerzitě Karlově č. P10 Lingvistika”, sottoprogetto “Románské jazyky ve světle jazykových korpusů”.

2.1.1 Il suffisso *-mento*

Come vedremo più in avanti, le restrizioni inerenti al suffisso *-mento* hanno una notevole stabilità diacronica. Infatti, tutti i vincoli che di solito vengono descritti valgono anche per la lingua cinquecentesca.

Come messo in evidenza già da Thornton 1987–1988: 356–357 (cfr. anche Gaeta 2004: 327), la principale condizione fonologica richiede, per *-mento*, che le basi non siano monosillabiche, per cui lì dove il tema verbale che funge da base sia monosillabico (*dire, fare, andare, ecc.*), viene preferito un tema irregolare (sempre stando alla formulazione di Thornton 1987–1988: 357): *dicimento, facimento, andamento*.

È però molto probabile che l'esclusione di tali forme sia causata da un vincolo fonologico più generale, formulato, in termini della Teoria dell'Ottimalità (OT), da Burzio (2003), ma discusso (ovviamente, nell'ambito di un quadro teorico diverso) anche da Thornton (1987–1988: 203–210). La formulazione di Burzio prevede una gerarchia generale di due vincoli principali: la condizione metrica e la condizione morfologica (che riguarda la realizzazione "segmentale" delle forme). Secondo Burzio (2003: 75), la condizione metrica è più forte di quella che richiede il mantenimento del materiale morfologico della base anche nel derivato. Pertanto le forme ipotetiche come **dimento, *famento, *vamento* sono escluse perché le forme – che secondo Thornton dovrebbero teoricamente fungere da basi (essendo identiche all'imperativo) – quali *di-, fa-, va-* sono accentate, così come è accentato anche il suffisso *-mento*. Lo "scontro accentuale" che si sarebbe creato (**váménto*) si evita con maggior rigore rispetto ad una modificazione morfologica della base.

Per quel che riguarda la serie di restrizioni morfologiche, occorre soffermarsi sulla questione dell'appartenenza ad una data classe flessiva. A. M. Thornton (1987–1988: 332), in base al suo campione, osserva come la cooccorrenza "*tra un dato suffisso e una data coniugazione*" non sia particolarmente evidente, ma al contempo ammette, in accordo con Scalise (1994: 115), un legame stretto tra la coniugazione in *-ire* e il suffisso *-mento*. Gaeta (2004: 327), invece, sostiene che "... *non sussistono relazioni dirette che vincolano uno dei due suffissi ad una classe flessiva*", venendo così a contraddirsi proprio quando osserva come i verbi parasintetici della III coniugazione richiedano il suffisso *-mento*, il che, afferma Gaeta (2004: 329) "[*mostra*] quindi una correlazione stretta tra classe flessiva e selezione del suffisso." Agli esempi adottati dallo stesso studioso (*intestardimento, intiepidimento, involgarimento*), possiamo aggiungerne altri, tratti dal corpus *La Repubblica* o da Internet (per mezzo del motore di ricerca Google), quali *rincretinimento* (vs. **rincretinizione*) o *incafonimento* (vs. **incafonizione*).

Quanto alle altre restrizioni, è ben noto il fatto che certi suffissi presenti nella base verbale (come i verbi in *-eggiare*) favoriscano l'aggiunta del suffisso *-mento* (cfr. la tabella 5 di Gaeta 2004: 331). Altrettanto nota è anche la selezione di *-mento* con le basi verbali prefissate di "tradizione popolare" (*ad-, in-, s-*). Infine, *-mento* dimostra anche una preferenza per basi verbali semplici, anche se proprio in questo caso, la concorrenza con il suffisso *-zione* (evidente anche dal punto di vista quantitativo, cfr. la tabella 6 di Gaeta 2004: 331) può essere rivelatrice di un'istruzione semantica leggermente diversa.

2.1.2 Il suffisso *-zione*

Invece, per il suffisso *-zione*, il discorso delle restrizioni risulta, dal punto di vista diacronico, alquanto diverso.

Al livello fonologico, il suffisso *-zione* viene evitato laddove la base verbale finisce in una sequenza formata dall'affricata sorda [ts] e una vocale, come in *aguzzare* → **aguzzazione*, *deprezzare* → **deprezzazione* (cfr. Thornton 1987–1988: 356; Gaeta 2004: 327). A differenza della restrizione fonologica del suffisso *-mento*, si tratta in questo caso piuttosto di un vincolo eufonico.

Per ciò che riguarda la morfologia, la presenza di certe sequenze morfologiche favorisce la selezione di *-zione*, come avviene per verbi in *-izzare* e *-ificare*. Se questi verbi rappresentano così l'*input* principale per il suffisso *-zione*, la loro presenza nella lingua cinquecentesca è ancora molto limitata (cfr. *infra*).

Inoltre, anche le basi prefissate (per mezzo di prefissi “colti”) mostrano una preferenza per *-zione*: anche se, in realtà, la presenza – tanto per fare un esempio – del prefisso *de-* in verbi quali *defonologizzare*, *degassificare*, non può certo dirsi rilevante per la selezione del suffisso *-zione*; più importante è, ovviamente, la sequenza finale della base verbale. Inoltre, il presunto influsso del prefisso “colto” *de-* sulla selezione del suffisso (cfr. Gaeta 2004: 329) è certamente da escludere nelle formazioni come *deperimento* e *denutrimiento*, in cui, come abbiamo già avuto modo di vedere, il suffisso *-zione* viene evitato per l'appartenenza del verbo di base alla III coniugazione. E infine lo stesso discorso varrebbe anche per le coppie come *decomposizione/decomponimento*, in cui la diversa selezione rispecchia senz'altro la differenziazione semantica che abbiamo già visto.

Come si vede, i due suffissi non possono considerarsi concorrenti nella misura in cui la loro applicazione viene regolata da una serie di restrizioni. Soltanto laddove la base ammette entrambi i suffissi, la questione della concorrenza può essere posta; ma nella maggior parte dei casi (di cui non possiamo, *faute de place*, occuparci), le formazioni quali *integrazione/integramiento* non sono del tutto concorrenti dal momento che il derivato in *-mento* permette solo una lettura processuale, mentre quella in *-zione* esibisce tipicamente anche una o più delle estensioni semantiche.

2.2 L'aspetto quantitativo e la nozione di produttività

Come abbiamo già premesso, per l'aspetto quantitativo si intende la *frequenza* con cui le formazioni in questione occorrono. Tale frequenza, che spesso viene identificata con la nozione di *produttività*, viene definita sia in base all'informazione lessicografica, sia in base ad un corpus testuale.

2.2.1 La *type frequency*

Il quadro che si ottiene basandosi sui dizionari corrisponde alla produttività intesa come *type frequency*, cioè come frequenza di lemma (cfr. Bauer 2001: 47). In tal modo, il processo produttivo può essere definito in base al numero di parole nuove create in un dato periodo: “*The more productive a morphological process is, the more coinages that*

occur created by that morphological process in a given time period. (...) the productivity of the process can be equated with the average number of new words (...) that are used in the language within a specified time period." (Bauer 2001: 156).

Come è risaputo, tale indagine, limitata a quanto registrato dai dizionari e applicata a processi intuitivamente produttivi perde un importante aspetto, cioè la potenzialità o, meglio, la probabilità di trovare o di creare una parola completamente nuova, ancora non registrata dai dizionari, eppure del tutto trasparente e correttamente formata dal punto di vista delle RFP (sulle limitazioni e sui vantaggi della ricerca basata sui dizionari cfr. Bauer 2001: 156–161; Gaeta – Ricca 2003: 63–65).

2.2.2 La token frequency

L'approccio basato su corpora elettronici richiede l'introduzione di una seconda dimensione che è quella associata alla *token frequency*. Infatti, in un corpus una data parola (*lemma* ovvero *tipo*) può occorrere varie volte e la frequenza di occorrenze diventa un fattore importante per due motivi. Il primo, di ordine psicolinguistico, è che le parole ad alta frequenza tendono ad essere le parole lessicalizzate, immagazzinate nel lessico mentale, molto spesso anche poco trasparenti che devono essere apprese come tali senza ricorrere ad una regola di formazione di parole. Invece, le parole a bassa frequenza sono di solito morfologicamente e semanticamente trasparenti e possono, di conseguenza, essere analizzate per mezzo di una RFP, in base all'istruzione semantica della regola stessa (cfr. Baayen 1992). Il secondo motivo è di ordine matematico: il numero di occorrenze è, ovviamente, dato dalla dimensione del corpus in questione: con la progressiva crescita della dimensione del corpus, il numero di occorrenze tenderà necessariamente ad aumentare (cfr. Baayen 1992: 113).

In un'indagine basata su un corpus, dobbiamo dunque tener conto di almeno tre variabili: il numero di lemmi/tipi (*type frequency*, d'ora in poi V) e il loro numero di occorrenze (*token frequency*, d'ora in poi N) che, se calcolati per un dato affisso ad esempio, corrispondono a quello che Baayen 1992: 113 chiama *item sample*; è la dimensione del corpus che rappresenta il campione di base (*frame sample*, d'ora in poi F). Il rapporto tra queste variabili è il seguente: "... *the values of N and V , as calculated from the item sample, depend on the size of the frame sample. For larger frame samples, larger values of N and V are to be expected for the item sample.*" (Baayen 1992: 113).

2.2.3 Lo spettro di frequenza e l'accrescimento del lessico

Il rapporto tra queste tre variabili può essere colto per mezzo di tre strutture che possono essere create nel programma *zipfR*, un package da usare all'interno del software statistico *R* (cfr. Baayen 2008 per un'introduzione generale alla statistica lessicale in *R*).

La prima struttura è lo *spettro di frequenza* (*frequency spectrum*): si tratta di una lista in cui i lemmi sono ordinati a seconda della loro frequenza (d'ora in poi m) partendo dagli *hapax legomena* (cfr. Baroni 2009; Baayen 2008: 222–236). In tal modo, si rappresentano, per così dire, le frequenze delle frequenze ("frequencies of frequencies", cfr. Baayen 2008: 230): si visualizza così la frequenza di tutti i lemmi con la frequenza 1 ($m = 1$, ovvero V_1), di tutti i lemmi che occorrono nel corpus due volte ($m = 2$, ovvero V_2), e così via di seguito fino all'ultima frequenza rappresentata dalla parola più

frequente di tutte. Lo spettro di frequenza è particolarmente utile perché ci permette di vedere il numero di lemmi come funzione della loro frequenza $V(m)$.

La seconda struttura è la *curva di accrescimento del lessico* (vocabulary growth curve): come si è visto, il numero di occorrenze è dato dalla dimensione del corpus; così il numero di lemmi dipende indirettamente dal numero di occorrenze, come dice Baayen (1992: 113): “... for some fixed morphological process, V can be viewed as a function of N : for increasing numbers of tokens in the item sample, obtained by increasing the frame sample, V will also increase.” Si può rappresentare questa funzione per mezzo di un grafico in cui sull’asse x si colloca il numero di occorrenze N , e sull’asse y il numero di lemmi V . La curva che così viene creata si chiama, appunto, *curva di accrescimento del lessico* (cfr. Baayen 1992: 113; 2008: 222).

Di solito, si lavora con tre tipi di curva: empirica, interpolata ed estrapolata. Soprattutto la curva *estrapolata* è di massima importanza per lo scopo del presente lavoro. Infatti, l’estrapolazione binomiale consente di oltrepassare il numero reale di N dato dal campione. L’interesse di tale procedimento sta nel fatto che avendo un campione limitato non si possono azzardare delle ipotesi sul totale della *popolazione*, ovvero sul numero possibile di tutti i lemmi a cui una data regola può dar luogo (cfr. Baroni – Evert 2006). Inoltre, avendo un campione limitato e, in molti casi, di dimensioni diverse per ogni processo della RFP, non si può procedere al mutuo confronto dei processi, perché il confronto non sarebbe completo dal momento che le due curve non sarebbero quantitativamente paragonabili.

La terza struttura è il noto indice P che coglie il *ritmo di accrescimento del lessico* (vocabulary growth rate): il valore P , calcolato nel modo seguente $P = V_1 / N$ (cioè il numero dei lemmi con la frequenza 1, gli *hapax legomena*, diviso per il totale delle occorrenze del tipo morfologico in questione), indica il *ritmo* con cui il lessico cresce (cfr. Baayen 1992: 115; 2008: 222). Tale ritmo viene inteso come la probabilità di trovare un lemma nuovo in relazione al progressivo aumento della dimensione del corpus: “...it expresses in a very real sense the probability that new types will be encountered when the item sample is increased.” (Baayen 1992: 115).

3. Corpus ed elaborazione dei dati

I dati – le formazioni con il suffisso *-mento* e *-zione* – che presenteremo e analizzeremo sono stati estratti dal corpus LIZ 4.0, una raccolta di testi della letteratura italiana che va dal Duecento alle soglie del Novecento; il corpus comprende 1000 testi di circa 40 mil. di *tokens*.

All’interno del LIZ 4.0, ci siamo avvalsi della periodizzazione prestabilita limitando la selezione al sottocorpus del Cinquecento (di circa 10 mil. di occorrenze, 231 testi). In seguito, abbiamo estratto le liste di forme terminanti in sequenze che corrispondono ai due suffissi presi in esame; nella ricerca “sequenziale” abbiamo inserito tutte le forme che i due suffissi potevano assumere prestando particolare attenzione alle varianti ortografiche: ad esempio, la lista del suffisso *-zione* conteneva le forme uscenti in *-zione/-zioni/-zion, -sione/-sioni/-sion, -tione/-tioni/-tion-*; la lista così ottenuta è stata ulteriormente ricontrollata con una ricerca complementare sulla sequenza *-ione/-ioni/-ion*. In seguito, abbiamo proceduto alla lemmatizzazione e all’eliminazione dei tipi che non appartenevano alla RFP in questione (per la descrizione dei criteri cfr. Štichauer 2009: 72–74).

Il profilo quantitativo che abbiamo così ottenuto è riassunto nella tabella 1 dove vengono presentate tutte le variabili introdotte nel paragrafo precedente.

Tabella 1. I valori di V , N e V_1 per i suffissi -mento e -zione corrispondenti al corpus del Cinquecento ($F = 10604452$)

SUFFISSO	V	N	V_1
<i>-mento</i>	584	14028	281
<i>-zione</i>	722	32031	194

Il quadro è parzialmente diverso da quelli che si riscontrano nei secoli precedenti (cfr. Štichauer 2009). Infatti, il suffisso *-zione* esibisce più o meno le stesse proporzioni: il numero dei lemmi è più alto, ma gli *hapax legomena* rappresentano sempre circa un quarto del numero totale di V , mentre il suffisso *-mento* mantiene sempre alta la proporzione degli *hapax legomena*: quasi il 50%. Anche la frequenza delle occorrenze è importante: il valore di N è doppio rispetto a *-mento*.

Questa situazione è già tipica anche dell'italiano odierno in cui il suffisso *-zione* possiede un maggior numero di lemmi rispetto a *-mento* ed esibisce anche un'elevatissima *token frequency*, ma rimane inferiore per quanto riguarda il numero degli *hapax legomena* e la loro proporzione con il numero di lemmi. In effetti, si può affermare che la situazione cinquecentesca preannunci in qualche modo la situazione odierna anche se le frequenze relative dei due suffissi presentano una differenza notevole che si riassume nella tabella 2 (per l'italiano contemporaneo cfr. Gaeta – Ricca 2003).

Tabella 2. La frequenza relativa delle occorrenze (*token frequency*) dei suffissi *-mento* e *-zione* nel corpus del Cinquecento e nel corpus della *Stampa* (secondo Gaeta – Ricca 2003)

SUFFISSO	Corpus del '500 ($F = 10$ mil. di <i>tokens</i>)	Corpus <i>La Stampa</i> ($F = 75$ mil. di <i>tokens</i>)
	<i>frequenza relativa</i>	<i>frequenza relativa</i>
<i>-mento</i>	1,32 ‰	3,4 ‰
<i>-zione</i>	3,02 ‰	13,9 ‰

La frequenza relativa normalizzata, che ci permette, appunto, di paragonare due corpora di dimensioni variabili, indica che il suffisso *-zione* subisce, nell'arco del tempo che va dal Cinquecento ai giorni nostri, un incremento massiccio della sua *token frequency* mantenendo però la stessa proporzione del numero dei lemmi e degli *hapax legomena*, come si evince dalla tabella 3.

Tabella 3. Il numero dei lemmi (V) dei suffissi *-mento* e *-zione* nel corpus del Cinquecento e nel corpus della *Stampa* (secondo Gaeta – Ricca 2003)

SUFFISSO	Corpus del '500 (F = 10 mil. di tokens)	Corpus <i>La Stampa</i> (F = 75 mil. di tokens)
	V / V ₁	V / V ₁
<i>-mento</i>	584 / 281	1403 / 402
<i>-zione</i>	722 / 194	2363 / 486

4. Le formazioni a bassa frequenza – alcuni *hapax legomena*

Gli *hapax legomena* (V₁) rappresentano, come abbiamo già anticipato, un gruppo di formazioni molto importante, e non soltanto perché coprono dal 30 al 50% di tutti i lemmi (su questo aspetto, inerente ai corpora in generale, cfr. Baroni 2009), ma soprattutto perché si presume che siano un buon indizio della produttività di un dato suffisso nel senso che abbiamo già introdotto sopra in 2.2.3.

4.1 Il suffisso *-mento*

Tra le formazioni che nel corpus del Cinquecento occorrono una sola volta prevalgono naturalmente quelle che presentano una semantica del tutto trasparente; molto spesso, infatti, servono soltanto alla nominalizzazione del verbo precedente in un contesto più ampio.

Tanto per fare qualche esempio – perché ovviamente non si possono percorrere tutti i 281 *hapax* presenti nel corpus – possiamo soffermarci su tre formazioni interessanti: *aloppiamento*, *consequimento* e *guarimento*.

Alloppiamento, che ricorre in un passo della *Istoria delle Indie* di Ramusio, è il nome derivato dal verbo *aloppiare*, attestato già alla fine del Trecento con il significato di “addormentare con una sostanza oppiata” (cfr. TLIO, s.v. *aloppiare*). La trasparenza della formazione si vede nell’esempio (1).

- (1) (...) *perché, s’egli nol disse avanti, vuol che lo lascino stare a quel modo fin che passi quello **aloppiamento**, e che si digerisca il vino e il fumo.* (Giovanni Battista Ramusio, *Della naturale e generale istoria dell’Indie occidentali* di G. F. d’Oviedo, Lib. 5, Cap. 2, 3).

Anche *consequimento* è una formazione del tutto regolare, derivata dal verbo *consequire* in un contesto che richiede, naturalmente, la nominalizzazione del verbo, come si vede nell’esempio (2).

- (2) *Il medesimo è in noi, a cui la natura ha donato un appetito di bramar ogni cosa che buona ci paia (...), il che è che, secondo i peripatetici, l’appetito concupiscibile ha anche a noi fatto cortese dono d’un appetito, col quale ci sforziamo di far contesa a chi vietar ci volesse il **consequimento** del bene, (...)* (Matteo Bandello, *Parte 1, Nov. 10, Dedicata*).

Ci sono anche due casi di formazioni che sono in (debole) concorrenza con lemmi più frequenti: *crocifiggimento* / *crocifissione*, *guarimento* / *guarigione*. Tralasciando *crocifig-*

gimento, attestato in Aretino, prendiamo *guarimento* che si può vedere contestualizzato nell'esempio (3).

- (3) *Tutto da que' dolenti genitori agevolmente il maestro ottenuto, dispose l'opera sua al **guarimento** della fanciulla, (...)* (Scipione Bargagli, *Trattenimenti*, Pt. 1, Novella 1).

Anche se è molto difficile, in prospettiva diacronica, stabilire la differenza semantica con *guarigione*, tanto più che anche *guarigione*, nel corpus del Cinquecento, è una formazione a bassa frequenza, *guarimento* sembra corrispondere ad una nominalizzazione "dinamica", in cui l'azione del guarire, a cui il *maestro* si accinge, viene rappresentata come un processo, come un'operazione a cui la *fanciulla* viene sottoposta.

4.2 Il suffisso -zione

Gli *hapax legomena* tra i sostantivi con il suffisso *-zione* esibiscono, invece, una struttura semantica e morfologica un po' diversa (sulla quale ritorneremo in 6).

Infatti, siamo di fronte a formazioni che, nella maggior parte dei casi, trovano la propria origine in una voce latina (ad esempio, *inculcazione*). Rientrano, però, nel sistema linguistico assieme alla base verbale da cui sono derivate per cui sarebbe erroneo eliminarle dalle liste di lemmi su cui il computo della produttività si basa (per la lista completa dei 194 *hapax* si veda Appendice, in Štichauer 2009: 9.4.2).

Inoltre, ci sono anche alcune formazioni concorrenti con quelle con il suffisso *-mento*: *delirazione* e *dileguazione* si trovano, nel Cinquecento, accanto a *deliramento* (anch'esso *hapax legomenon*) e *dileguamento* (quest'ultimo un po' più frequente, con 3 occorrenze). Lo stesso discorso varrebbe anche per *accecazione* che si riscontra, per la prima volta, nel Quattrocento (già con 4 occorrenze, però), mentre nel Trecento troviamo ancora *accecamiento*.

Invece, la formazione *inspessazione*, derivata dal verbo *inspessare*, attestato nell'opera di Giordano Bruno, è ormai una parola regolarmente derivata per mezzo del suffisso *-zione*. La formazione contestualizzata si vede nell'esempio (4).

- (4) (...) *e non monteranno in alto né scenderanno al basso, sin tanto che o acquistando per apposizion di parti o per **inspessazione** dal freddo gravità maggiore, (...)* (Giordano Bruno, *De l'infinito universo e mondi*, Dialogo 5, 48).

5. La produttività dei due suffissi

Avendo presentato tutti i dati relativi ai due suffissi, possiamo procedere al confronto della loro produttività. Affinché il confronto sia quantitativamente corretto, si deve sempre paragonare a parità di *N*, cioè a parità del numero delle occorrenze di un dato suffisso. Tuttavia, come si è già visto, i due suffissi presi qui in esame si differenziano anche per il valore di *N*, perciò è importante "unificare" il numero di occorrenze. Uno dei modi di unificarlo è quello di *estrapolare* il valore *N* di un processo morfologico al valore più alto di un altro processo morfologico. In tal modo, potremmo pensare di estrapolare *N* (14028) del suffisso *-mento* al valore *N* (32031) del suffisso *-zione*.

Per farlo, bisogna far ricorso ai modelli parametrici di distribuzione delle frequenze (*Large-Numbers-of-Rare-Events*, modelli LNRE). All'interno dello *zipfR*, si hanno a disposizione tre modelli: *Generalized Inverse Gauss Poisson*, GIGP (elaborato da Baayen 2001: cap. 4); *Zipf-Mandelbrot*, ZM; e *finite Zipf-Mandelbrot*, fZM (entrambi elaborati da Evert 2004; cfr. anche Baroni – Evert 2006). Ciascuno dei tre modelli offre un calcolo diverso dei parametri basato sullo spettro di frequenza di partenza; i valori stimati sono di conseguenza diversi per ciascun modello, ed è quindi importante valutare la qualità dell'estrapolazione (cfr. Evert – Baroni 2006).

Per il campione preso qui in esame, è stato adottato (in base a vari risultati, cfr. Štichauer 2011) il modello ZM. Presenteremo qui di seguito due risultati a parità di $N = 15000$ (che è quello vicino al valore empirico del suffisso *-mento*) e $N = 30000$ (quello, invece, del suffisso *-zione*).

5.1 La curva di accrescimento del lessico

Come già introdotto sopra, la curva di accrescimento del lessico si rappresenta mediante un grafico in cui sull'asse x si colloca il numero di occorrenze N , e sull'asse y il numero di lemmi stimato in base al modello $E(V)$ (dove E sta per "estimated").

Le curve di accrescimento per i due suffissi si possono vedere in Figura 1, dove la linea verticale intermittente indica la soglia – il valore di N empirico del suffisso *-mento* – oltre la quale si *estrapola*.

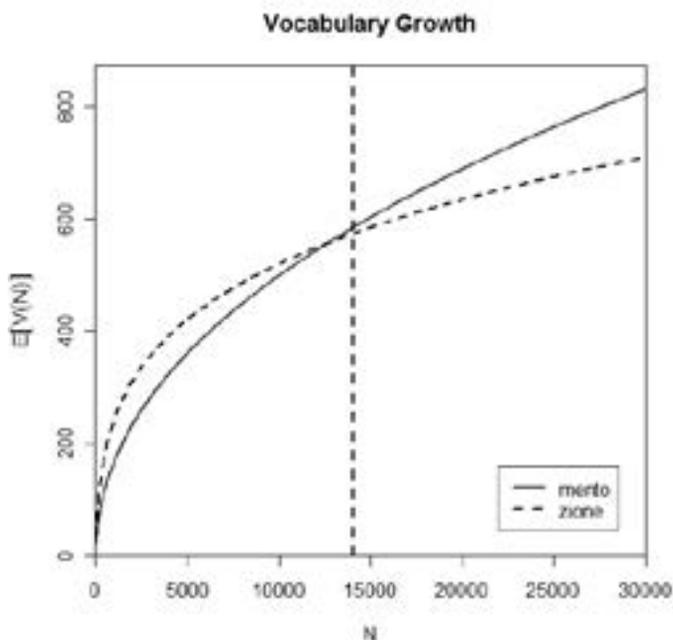


Figura 1. Le curve di accrescimento stimate in base al modello ZM per i suffissi *-mento* e *-zione* nel corpus del Cinquecento a parità di $N = 30000$

Le curve indicano che entrambi i suffissi sono produttivi nel senso che il numero dei lemmi, con il progressivo aumento della dimensione del corpus, tende a crescere. Tuttavia, l'accrescimento del suffisso *-mento* è, dopo un certo intervallo, più veloce di quanto lo sia quello di *-zione*. Ciò significa che dopo un rapido incremento dei lemmi, il suffisso *-zione* tenderà a rallentare questo ritmo.

5.2 Il valore P a parità di N

Questa tendenza la si vede bene anche nei valori di $E(V)$, $E(V_1)$ e P stimati a parità di due N già introdotti, cfr. la tabella 4.

Tabella 4. La produttività P dei suffissi *-mento* e *-zione* nel Cinquecento, stimata in base al modello ZM a parità di $N = 15000$ e $N = 30000$ (i valori di $E(V)$ a $E(V_1)$ sono arrotondati)

SUFFISSO	N	E(V)	E(V ₁)	P (E(V ₁)/N)
<i>-mento</i>	15000	602	280	0,0186
<i>-zione</i>	15000	584	167	0,0111
<i>-mento</i>	30000	830	383	0,0127
<i>-zione</i>	30000	709	193	0,0064

Come si vede, anche i valori di P confermano quanto già affermato: sebbene il suffisso *-zione* possieda, dato il corpus del Cinquecento, un maggior numero di lemmi, il ritmo di accrescimento è un po' meno rapido rispetto a *-mento*.

6. La struttura morfologica dei suffissati con *-zione* cinquecenteschi

Naturalmente, i dati numerici sono solamente indicativi di certe caratteristiche su cui ci soffermeremo brevemente in quest'ultima sezione prestando maggiore attenzione al suffisso *-zione*.

Infatti, nel ventaglio delle formazioni con il suffisso *-zione*, che vanno da quelle più frequenti agli *hapax legomena*, mancano alcune strutture interne che invece si trovano nella lingua contemporanea e che sono già state formate per mezzo di un procedimento *disponibile* (nel senso di Corbin 1987).

Queste strutture, che rappresentano il maggiore *input* per il suffisso *-zione*, sono i verbi con i suffissi causativi *-izzare* e *-ificare* (cfr. anche Gaeta – Ricca 2003; 2006). Tra tutti i 722 lemmi riscontrati nel corpus del Cinquecento, solo pochissimi rientrano tra le formazioni con *-izzazione* o *-ificazione*. Si tratta, per l'esattezza, di 15 sostantivi in *-ificazione* (*amplificazione*, *bonificazione*, *falsificazione*, *fortificazione*, *gratificazione*, *modificazione*, *mollificazione*, *pacificazione*, *prolificazione*, *purificazione*, *ratificazione*, *reedificazione*, *retificazione*, *santificazione*, *vivificazione*) e di una sola formazione in *-izzazione* (*particolarizzazione*).

Tale scarsa “rappresentanza” quantitativa è sicuramente dovuta – nell’estremo caso delle formazioni in *-izzazione* – all’assenza dei verbi con il suffisso *-izzare*. Infatti, come si evince dalla testimonianza lessicografica che si sviluppa soltanto dopo il periodo cinquecentesco, il suffisso verbale causativo *-izzare* diventa un procedimento *disponibile* solo nei periodi successivi. Un’ottima immagine in merito la si può ricavare dal confronto del numero dei verbi con *-izzare* (e il corrispondente numero dei sostantivi derivati con *-zione*) attraverso tre maggiori dizionari: il *Vocabolario degli Accademici della Crusca* (1612), il *Tommaseo* (1857–1879) e *Dizionario Italiano Sabatini-Coletti* (DISC 2004). I numeri sono senz’altro interessanti e non necessitano di commenti particolari, cfr. la tabella 5.

Tabella 5. Il numero dei verbi in *-izzare* e dei sostantivi in *-izzazione* in tre maggiori dizionari italiani

DIZIONARIO	Verbi in <i>-izzare</i>	Nomi in <i>-izzazione</i>
<i>Vocabolario della Crusca</i> (1612)	55	0
<i>Il Tommaseo</i> (1857–1879)	288	35
<i>DISC</i> (2004)	636	522

7. Conclusioni

Come si evince dalla tabella 5, il processo per cui il suffisso *-zione* è diventato un procedimento *disponibile* (ed anche altamente *redditizio*, sempre nel senso della *rentabilité* di Corbin 1987: 177) è lungo e, soprattutto, alquanto recente. Di fatto, si è dovuto aspettare che un altro processo diventasse, a sua volta, produttivo. In effetti, tutte le formazioni cinquecentesche in *-ificare* e *-izzare* non sono altro che *prestiti diretti* dal latino. In tal modo, il vasto gruppo dei sostantivi in *-zione* del Cinquecento si è insediato nella lingua italiana non tanto grazie alle proprie risorse morfologiche di formazione di parole, quanto piuttosto mediante un massiccio processo di *borrowing* morfologico (nel senso di Matras 2009: 209–212).

Questo percorso dal *prestito* al *processo autonomo* di formazione delle parole indica che uno dei cambiamenti interessanti nel sistema morfologico dell’italiano sarebbe da ricercare nel periodo che va dalla fine del Cinquecento alle soglie del ventunesimo secolo. Si apre così uno scenario nuovo ed un campo di ricerca tutto da esplorare.

RIFERIMENTI BIBLIOGRAFICI

- Baayen, H. (1992): Quantitative aspects of morphological productivity. In: G. Booij – J. van Marle (eds.). *Yearbook of Morphology 1991*. Dordrecht: Kluwer, 1992, pp. 109–149.
- Baayen, H. (2001): *Word frequency distributions*. Dordrecht: Kluwer.
- Baayen, H. (2008): *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.
- Baroni, M. – Evert, S. (2006): The *zipfR* package for lexical statistics: A tutorial introduction. Disponibile su: <http://zipfr.r-forge.r-project.org/>.

- Baroni, M. (2009): Distributions in text. In: A. Lüdeling – M. Kytö (eds.), *Corpus Linguistics. An International Handbook*. Berlin: Mouton de Gruyter, vol. 2, article 37, pp. 803–822.
- Bauer, L. (2001): *Morphological Productivity*. Cambridge: Cambridge University Press.
- Burzio, L. (2003): Output-to-Output Faithfulness in Phonology: the Italian Connection. *Lingue e Linguaggio*, 1, 2003, pp. 69–104.
- Corbin, D. (1987): *Morphologie dérivationnelle et structuration du lexique*, 2 voll., Tübingen: Niemeyer.
- Evert, S. – Baroni, M. (2006): Testing the extrapolation quality of word frequency models. *Proceedings of Corpus Linguistics 2005*. Birmingham, UK.
- Evert, S. (2004): A simple LNRE model for random character sequences. *Proceedings of JADT 2004*, pp. 411–422.
- Gaeta, L. – Ricca, D. (2003): Frequency and productivity in Italian derivation: A comparison between corpus-based and lexicographical data. *Italian Journal of Linguistics / Rivista di Linguistica* 15, 1, pp. 63–98.
- Gaeta, L. – Ricca, D. (2006): Productivity in Italian word formation: A variable-corpus approach. *Linguistics* 44, 1, pp. 57–89.
- Gaeta, L. (2004): Nomi d'azione. In: M. Grossmann – F. Rainer (a cura di): *La formazione delle parole in italiano*. Tübingen: Niemeyer, 2004, pp. 314–351.
- Matras, Y. (2009): *Language Contact*. Cambridge: Cambridge University Press.
- Scalise, S. (1994): *Morfologia*. Bologna: il Mulino.
- Štichauer, P. (2009): *La produttività morfologica in diacronia: i suffissi -mento, -zione e -gione in italiano antico dal Duecento al Cinquecento*. Praha: Karolinum.
- Štichauer, P. (2011): The quantitative approach to morphological productivity in a diachronic perspective. In: A. Zeldes – A. Lüdeling (eds.): *Proceedings of Quantitative Investigations in Theoretical Linguistics 4*. Humboldt-Universität zu Berlin, pp. 87–89.
- Thornton, A. M. (1987–1988): *Sui nomina actionis in italiano*. Tesi di dottorato, Università degli Studi di Pisa, A. A. 1987–1988.

Pavel Štichauer

Istituto di Studi Romanzi, Università Carlo di Praga

nám. Jana Palacha 2, 116 38 Praha 1

pavel.stichauer@ff.cuni.cz

www.pavel-stichauer.cz