

ROSSELLA VARVARA

Misurare la produttività morfologica: i nomi d'azione nell'italiano del ventunesimo secolo

Nel presente contributo si intende offrire una discussione critica delle metodologie, mutate velocemente negli ultimi decenni, utilizzate per misurare la produttività morfologica. Queste verranno applicate nell'analisi della formazione di nomi d'azione nell'italiano contemporaneo. Attraverso tale studio si mostreranno le criticità della misura ad oggi più utilizzata, ovvero la produttività potenziale (*P*, Baayen 1992, 2009), e si illustrerà come è possibile superarle mediante l'utilizzo di modelli statistici LNRE (Baayen 2001; Evert 2004; Evert & Baroni 2007).

Parole chiave: produttività morfologica, nomi d'azione, misure quantitative, modelli LNRE, morfologia derivazionale.

1. Introduzione

Nonostante siano passati poco più di due decenni dal convegno SLI tenutosi a Lugano nel 1991, le metodologie utilizzate in linguistica si sono evolute rapidamente. La disponibilità di grandi *corpora* testuali e di strumenti per la loro processazione ha reso sempre più diffuse analisi quantitative basate sui dati, permettendo una più rapida ed esatta disamina delle teorie proposte.

Nel campo della morfologia e, più nello specifico, della formazione delle parole, i *corpora* disponibili si sono dimostrati validi strumenti per l'osservazione di neologismi e nuove tendenze, offrendo un vasto campione di dati naturalistici.

Il contributo di Iacobini e Thornton al convegno di Lugano, pubblicato negli atti editi nel 1992, utilizzava gli strumenti all'epoca disponibili per osservare la produttività di diversi suffissi italiani nella formazione di nuove parole. Oggi è possibile rivedere le conclusioni a cui i due studiosi erano giunti e fare un raffronto non soltanto tra le

tendenze da loro individuate e quelle presenti nell'italiano degli ultimi due decenni, ma anche tra le metodologie impiegate.

In questo contributo si è scelto come caso studio la produttività dei nomi d'azione in italiano, osservando in un *corpus* di italiano contemporaneo (Paisà, Lyding *et al.* 2014) quale suffisso è più frequente per la formazione di questa classe di nomi.

Nella sezione 2 si discuterà il concetto di produttività morfologica da un punto di vista teorico e si illustreranno le principali misure proposte per la sua valutazione. Nella sezione 3 si introdurrà la classe dei suffissi analizzati e si riassumeranno i contributi precedenti che ne hanno osservato la produttività. Nella sezione 4 si applicheranno le misure quantitative descritte in §2 per calcolare la produttività dei nomi d'azione. Infine, nella sezione 5, si trarranno le conclusioni e si prospetteranno indirizzi futuri di ricerca.

2. *La produttività morfologica: definizioni e misure*

La produttività può essere definita come la capacità di un processo morfologico di essere usato sistematicamente per formare nuove parole e, in aggiunta, come la misura in cui ciò avviene (Plag 2006: 127). Una simile definizione è espressamente bipartita: da una parte indica la produttività come una caratteristica qualitativa che è o non è posseduta da un processo morfologico; dall'altra, come una proprietà scalare, che può essere riscontrata in maggiore o minore misura. Tale bipartizione segue il lungo dibattito teorico¹ su questa nozione che ha portato molti linguisti a proporre una suddivisione della nozione. Un simile approccio, difeso da Bauer (2001) e ripreso da molte definizioni seguenti (ad esempio, Plag 2006), può essere ricondotto a Corbin (1987), la quale individuava come *disponibilité* (tradotto in inglese da Carstairs-McCarthy, 1992, come *availability* e in italiano come *disponibilità* da Thornton, 2005) la possibilità di un suffisso di essere usato nella costruzione di un neologismo (Corbin 1987: 177) e con *rentabilité* (*profitability* in inglese e *rendimento* in italiano) l'effettivo utilizzo nella formazione di un ampio numero di lessemi.

¹ Per un *excursus* sulla storia e sulle diverse definizioni della nozione di produttività si rimanda a Bauer (2001, 2005).

Sin dalle prime trattazioni esplicite, si è utilizzato il numero di parole formate mediante un determinato processo per offrire un'idea intuitiva di produttività (Aronoff 1976: 36). Seguendo questo approccio, potremmo dire che un processo è *disponibile* in un determinato periodo storico se presenta almeno una nuova formazione e potremmo valutarne il *rendimento* in base al numero di formazioni prodotte. Tuttavia, ci possiamo aspettare che il numero di processi morfologici totalmente improduttivi sarà effettivamente molto basso e che è invece rilevante notare la differenza in termini di rendimento tra un suffisso che ha prodotto centinaia di nuove forme e uno che ne ha prodotte solamente due. Per questo motivo le ricerche sulla produttività si sono concentrate maggiormente sull'aspetto quantitativo, ovvero sul rendimento di un processo, piuttosto che sulla distinzione qualitativa.

Oltre al numero di lemmi formati da un affisso, una prima misura di produttività è stata proposta da Aronoff (1976). Essa corrisponde al rapporto tra il numero di parole effettivamente formate mediante un processo morfologico (*'actual words'*) e il numero di parole correttamente possibili (*'possible words'*). Più alto il valore di questo rapporto, più alta la produttività della regola morfologica. Aronoff lega il concetto di *parola possibile* al dominio di applicazione di una regola: le parole possibili sono quelle che si formerebbero dalle basi ammesse. Tuttavia, questa idea presenta diversi problemi. In primo luogo, non è facile definire i domini di applicazione di un processo morfologico. Molti sono i tipi di restrizioni in gioco (si veda, ad esempio, Rainer 2005 per una rassegna) e spesso è possibile osservare solamente delle preferenze, non delle divisioni nette tra domini. Di conseguenza, definire il numero di basi accettabili e quindi di parole possibili è arduo. Inoltre, non è chiaro il confine tra parole possibili e parole effettive: si potrebbero ad esempio contare come parole effettive quelle elencate in un dizionario, ma Aronoff stesso era ben conscio del fatto che nelle risorse lessicografiche i lemmi derivati da processi regolari spesso non sono riportati, data la loro facilità di interpretazione. Infine, come notato da Plag (2006: 122), questo indice non funziona correttamente per processi estremamente produttivi o totalmente improduttivi.

A causa di queste criticità, la nozione di parole possibili non è stata inclusa nelle formulazioni successive delle misure di produttività. L'attenzione è stata rivolta alle attestazioni effettive dei derivati di

un processo, utilizzando il numero dei lemmi prodotti o il numero di neologismi ed estraendo questi dati da risorse lessicografiche o da *corpora*.

Consideriamo in primo luogo le misure basate su dati provenienti da dizionari. Da queste opere è possibile estrarre, per l'affisso di interesse, il numero di formazioni (ovvero, il numero dei *types*) o dei neologismi formati in un dato periodo di tempo. Tuttavia nel primo caso, si otterrà una stima di quanto quel processo morfologico è stato produttivo in passato. È possibile, infatti, che una lingua conservi molte o alcune formazioni create in passato, senza che i parlanti utilizzino ancora quel processo per formarne di nuove. Il suffisso atono *-ita*, ad esempio, non è più produttivo in italiano (Gaeta 2004: 350), sebbene alcune sue formazioni siano ancora molto usate (*crescita, nascita, cernita*, ecc.). Il calcolo del numero di neologismi formati in un determinato periodo di tempo, laddove la risorsa includa la data di prima attestazione, costituisce una miglior valutazione della produttività. Plag (1999, cap. 5), ad esempio, sulla base dei dati estratti dall'*Oxford English Dictionary*, osserva che il suffisso inglese *-ize* ha dato vita a 284 nuovi verbi tra il 1900 e il 1985, fornendo prova della produttività di tale suffisso. Tuttavia, questa misura (e con essa anche la precedente) risente del campionamento effettuato dal lessicografo nella compilazione del dizionario. Spesso, come già accennato, in queste risorse vengono inseriti i termini percepiti come meno trasparenti, ovvero meno facilmente interpretabili. Parole la cui semantica è totalmente comprensibile mediante il significato dei morfemi che la compongono non avranno bisogno dell'aiuto del dizionario per essere comprese e per questo non verranno inserite.

Per ovviare a tali inconvenienti, negli ultimi decenni, a partire dal lavoro di Baayen del 1992, ci si è concentrati su misure quantitative basate su dati estratti da *corpora*.

Il calcolo del numero di *types* attestati nel *corpus*, chiamato *extent of use* (Baayen 1993), *type-frequency* (indicato con *V*) (Plag 2006), o *realized productivity* (Baayen 2009), similmente alla misura parallela calcolata su dizionari, presenta lo svantaggio di essere una misura della produttività passata. Questa resta comunque informativa, soprattutto dal punto di vista diacronico.

Così come per gli studi sui dizionari, anche nel caso dei *corpora* si è pensato di analizzare il numero di neologismi come misura di produttività. In questo caso non è possibile ricavare la data di prima attestazione e si è quindi giunti a utilizzare gli *hapax legomena* presenti in un *corpus* come approssimazione dei neologismi della lingua. È stato osservato (Baayen & Renouf 1996; Plag 2003) che, sebbene non tutti gli *hapax* siano neologismi, questi si concentrano principalmente tra le parole poco frequenti, in particolar modo tra gli *hapax*. A livello intuitivo potremmo dire che un neologismo non è ancora entrato ampiamente nell'uso linguistico e per questo, all'interno di un campione come un *corpus*, tenderà a occorrere una o poche volte.

Dato questo per assunto, una volta estratti gli *hapax* da un *corpus*, è possibile misurare la *hapax-conditioned degree of productivity* (Baayen 1993), chiamata anche *expanding productivity* (Baayen 2009).

Una misura ulteriore, simile alla precedente, è la cosiddetta *produttività potenziale* (*potential productivity* in Baayen 2009, o *category-conditioned degree of productivity* in Baayen 1993). Solitamente indicata con la lettera *P*, misura il rapporto tra il numero di *hapax* attestati con un determinato affisso (n_1^{aff}) il numero totale di occorrenze di tutte le forme dell'affisso (N^{aff}):

$$P = \frac{n_1^{aff}}{N^{aff}}$$

Idealmente questa misura prende in considerazione l'uso totale (passato e presente) di un determinato affisso e a ciò rapporta il numero di nuove forme. Ci offre una valutazione della probabilità di incontrare una nuova parola formata con il processo morfologico in questione all'interno di un testo.

Ad oggi questa è la misura più ampiamente utilizzata. Tuttavia, diverse critiche le sono state rivolte. Già il suo creatore (Baayen 1992: 117) aveva notato che essa dipende dalla grandezza del campione considerato: per valori maggiori di *N*, *P* decrescerà. Di conseguenza, se compariamo due processi con misure di *N* diverse, il suffisso meno frequente sarà favorito. Gaeta e Ricca (2002, 2006) hanno proposto di ovviare a questo problema calcolando *P* per valori uguali di *N*, invece che per il valore ottenuto in un intero *corpus*. In questo modo, però, essendo obbligati a considerare come *N* il numero di occorrenze

inferiore tra i valori dei due affissi, perdiamo il rapporto originale tra *hapax* e occorrenze totali per l'affisso più frequente.

È possibile risolvere questi problemi mediante l'utilizzo di modelli statistici chiamati LNRE (*Large Number of Rare Events*, Baayen 2001), in particolar modo con il modello Zipf-Mandelbrot finito (fZM, Evert 2004). Questi modelli permettono di calcolare P , sulla base della distribuzione osservata nel campione, per qualsiasi valore di N , e quindi di comparare processi morfologici con frequenze molto diverse o in *corpora* differenti. Più in generale, predicono la dimensione di un vocabolario (numero di *types*) per l'intera popolazione basandosi su un campione ridotto. Sono in grado, ad esempio, di predire il numero di parole che un bambino di cinque anni conosce, osservando un campione della sua scrittura. La loro accuratezza è stata testata per campioni di grandezze diverse (Baayen 2001; Evert 2004). Nel presente contributo non ci si può dilungare sulla loro descrizione matematica, per cui si rimanda ai riferimenti già citati. Si mostrerà l'uso di questi modelli nello studio della produttività dei nomi d'azione in italiano, dopo un *excursus* dei precedenti studi sul tema.

3. *Stato dell'arte: la produttività dei nomi d'azione*

In italiano, i suffissi *-ione*, *-aggio*, *-tura* e *-mento* possono essere utilizzati per formare nomi d'azione (anche detti nomi di evento), ovvero quella classe di nomi deverbali che esprimono il significato di *l'atto*, *l'evento*, *lo stato di -V*, dove V è il verbo dal quale sono derivati. Il loro uso e la loro alternanza è oggetto di numerosi studi (si veda ad esempio Gaeta 2004) che mirano a osservarne i domini di applicazione, le differenze di significato e la loro diversa produttività.

Tra i lavori dedicati a quest'ultimo punto, Thornton (1988) analizza la produttività di tali suffissi attraverso repertori di neologismi del ventesimo secolo. In particolar modo, considera i neologismi attestati nel *Dizionario di parole nuove 1964-1984* di Cortelazzo e Cardinale (1986) e nel *Dizionario moderno di parole che non si trovano negli altri dizionari* di A. Panzini (edizione dal 1905 al 1963). Thornton è ben consapevole del fatto che tali raccolte non siano comprensive di tutti i neologismi esistenti e che il campione analizzato dipenda in larga misura dalle scelte dei lessicografi. Cita infatti una ammissione di non esaustività di Cortelazzo e Cardinale:

se si dovessero comprendere in un repertorio di neologismi della lingua comune tutti i nuovi termini, che arricchiscono i linguaggi settoriali, il volume del dizionario si dilaterrebbe smisuratamente. Perciò è stata operata una necessaria selezione delle voci specialistiche, evitando di attingere alle fonti appropriate (...) e preferendo i mezzi più accessibili e di più larga circolazione, soprattutto i giornali, che effettuano già una discriminazione fra termini tecnici ormai di largo impiego e termini rimasti limitati agli specialisti di un determinato settore (p. iv, citato da Thornton 1988: 299)

Nonostante questo limite, l'autrice propone una interessante comparazione diacronica della produttività, grazie al diverso arco temporale coperto dalle risorse analizzate. Tra il 1942 e il 1986 osserva dunque che i suffissi *-zione* e *-aggio* sono in espansione, ovvero presentano una produttività progressiva, mentre i suffissi *-mento* e *-tura* rimangono stabili nel numero di neologismi attestati. Dal punto di vista qualitativo, Thornton nota che, in tale periodo, i derivati in *-tura* hanno perso una denotazione prettamente tecnica, formando termini appartenenti al linguaggio comune (*ammanicatura, incavolatura, incazzatura*). Il suffisso *-aggio* in entrambi i periodi risulta maggiormente utilizzato per termini tecnici e per basi straniere.

Iacobini e Thornton (1992) utilizzano la stessa procedura di spoglio dei repertori di neologismi e, per quanto riguarda i nomi d'azione, notano una forte produttività del suffisso *-zione*, accompagnata da una generale regressione degli altri suffissi. In entrambi gli studi, questo incremento dei nomi in *-zione* è legato alla parallela fortuna del suffisso verbale denominale *-izzare*, i cui derivati presentano una grande preferenza per ulteriori derivazioni in *-zione*. La sorte degli altri suffissi di azione, nonostante la stessa metodologia, appare diversa tra i due studi: Thornton (1988) utilizza un campione più ampio e individua così un valore diverso per i suffissi *-mento* e *-tura*, che in Iacobini e Thornton (1992) sembrano in regresso. Questo caso è esemplificativo di un problema che concerne qualsiasi tipo di misura della produttività, che sia basata sui *corpora* o su dizionari, ovvero la grandezza del campione.

Fiorentino (2008) calcola la produttività dei nomi d'azione basandosi sul numero di lemmi attestati in tre testi di diverse epoche (XIV, XVII e XX secolo). La sua analisi rivela che nel corso di questo periodo il suffisso *-zione* presenta un numero crescente di attestazioni, mentre i suffissi *-mento* e *-aggio* occorrono in numero decrescente.

Gaeta e Ricca (2002) adottano un approccio basato sul numero di *hapax* attestati in un *corpus* (nel loro caso, in un *corpus* giornalistico ottenuto da due annate, 1996-1997, del quotidiano La Stampa), proponendo una modifica alla misura di produttività di Baayen. Essi calcolano il valore di P per un numero fisso di *tokens* dell'affisso in esame, non considerando il numero totale di *tokens* nel *corpus*. La loro proposta si muove principalmente dalla consapevolezza che la misura P avvantaggia suffissi meno frequenti, data la presenza al denominatore del numero di occorrenze totali dell'affisso. In questo modo, tuttavia, la loro misura non considera la differenza di *tokens* e non permette comunque di comparare affissi con frequenze troppo diverse. I risultati da loro evidenziati mostrano una maggiore produttività dei suffissi *-(z)ione* e *-mento* rispetto agli altri. Gli autori sottolineano, inoltre, come questi sarebbero quasi ribaltati dalla misura standard P .

Conscio dei problemi precedentemente individuati riguardanti la misura proposta da Baayen e avendo bisogno di comparare *corpora* diacronici di diverse grandezze, Štichauer (2009) utilizza il modello ZM (Evert & Baroni 2007) per calcolare la misura P per uno stesso valore di N . Come già detto nella sezione precedente, basandosi sulla distribuzione dei dati osservata nel *corpus*, il modello predice il valore di P per un numero diverso di N . Vale a dire, ipotizza quale sarebbe la produttività di un affisso per un certo numero di *tokens*, osservando le proporzioni effettivamente attestate. Štichauer applica questa misura allo studio diacronico della produttività dei suffissi *-mento* e *-zione*, osservandone l'evoluzione tra il XIII e il XVI secolo. Il suffisso *-mento* appare costante nella sua applicazione nei diversi secoli, mentre il suffisso *-zione* mostra un aumento di produttività, con un picco nel XV secolo.

4. La produttività dei nomi d'azione nell'italiano contemporaneo

Gli studi sulla produttività dei nomi d'azione sono numerosi e, come si è visto, adottano metodologie diverse. In questa sezione si vuole mostrare la validità delle misure basate su modelli LNRE, applicandole allo studio sincronico della produttività dei suffissi d'azione e comparandole alla tradizionale misura P . Quest'ultima spesso fornisce dati controintuitivi e, nel caso in oggetto, stravolge i risultati di qualsiasi altra misura proposta. La versione di P basata sui LNRE, in-

vece, pur mantenendo lo stesso impianto teorico, non tende a favorire i processi morfologici meno frequenti. A differenza della proposta di Gaeta e Ricca (2002), invece, può essere applicata per la comparazione di affissi di frequenze molto diverse.

I suffissi presi in considerazione sono *-tura*, *-aggio*, *-mento* e *-zione* e lo scopo che ci si prefigge è quello di osservare quale suffisso è preferito nell'italiano contemporaneo per la formazione di nuovi nomi d'azione. Seguendo la letteratura precedente, si considereranno gli *hapax legomena* come stime approssimative dei neologismi e si utilizzerà il valore della loro frequenza nel calcolo di diverse misure di produttività.

Per l'estrazione di questi dati si è ricorso a un *corpus* di testi tratti dal web, il *corpus* Paisà (Lyding *et al.* 2014). I documenti sono stati collezionati tra il settembre e l'ottobre del 2010 per un totale di circa 250 milioni di *token*. Il linguaggio del web presenta diverse caratteristiche proprie del parlato (Berruto 2005), come ad esempio un'organizzazione delle informazioni meno gerarchizzata e meno coesa (Rossi 2010) o scarsa pianificazione e maggiore immediatezza (Bazzanella 2003), pur presentando una pluralità di forme espressive e di stili. Date queste peculiarità, ci possiamo aspettare che in un contesto meno controllato rispetto allo scritto tradizionale i parlanti utilizzino con maggiore libertà nuove formazioni, con una minore preoccupazione sulla loro correttezza.

Dalle liste di frequenza del *corpus* sono state estratte automaticamente tutte le occorrenze delle parole terminanti con i suffissi presi in considerazione. Prima delle analisi effettive, è stato svolto uno spoglio manuale dei lemmi ottenuti automaticamente. Diversi casi estrapolati non erano interpretabili come istanze dei processi in oggetto e per questo non sono stati presi in considerazione. Si sono eliminati in primo luogo i molti refusi presenti soprattutto tra le parole che occorrono solo una volta (ad esempio, *dichiariazione*, *dichiazazione*, *dichiazione*, *dichirazione*, *dichiuarazione*, *diciarazione*). Oltre a ciò, si è seguito quanto dettagliatamente proposto in Gaeta e Ricca (2002: 233 e seguenti) eliminando così:

- le parole non morfologicamente complesse o divenute opache semanticamente (*elemento*, *cemento*, *frumento*, *documento*, *stazione*, *temperatura*, *cultura*, etc.). Questo genere di lemmi presenta un'alta frequenza e avrebbe influenzato marcatamente i risultati.

Seguendo Gaeta e Ricca (2002), si sono esclusi i termini in cui non era più evidente un rapporto con la base e si sono invece considerati i casi polisemici, per i quali uno dei significati non era più azionale (es. *armamento, cucitura*).

- le parole derivate da basi non verbali (es. *formaggio, linguaggio, paesaggio, letteratura*). Questo problema si pone soprattutto per il suffisso *-aggio*, il quale può essere applicato anche a basi nominali.
- derivati da basi complesse in cui il processo morfologico in esame risultava essere anteriore. Soprattutto nei casi in cui era presente anche un processo di prefissazione si è dovuta prestare attenzione a quale procedimento era avvenuto prima. Questa scelta è probabilmente la più soggettiva fin qui applicata e avrebbe meritato un'analisi dettagliata per ogni singolo caso, purtroppo impossibile per il vasto campione selezionato. In generale, si sono inclusi solo quei termini in cui la prefissazione sembrava più chiaramente precedere la suffissazione (es. *rivascolarizzazione, ricoprimento, infornamento, sbendaggio*).

Va notato, infine, che il controllo manuale dei dati si è rivelato particolarmente arduo proprio nel caso degli *hapax*. Molti termini che a un primo sguardo sembravano insensati, frutto di un errore, si sono rivelati formazioni accettabili una volta che si è consultato il contesto d'uso nel *corpus*. Il termine *morulazione*, ad esempio, non è di immediata interpretazione e poteva essere eliminato se considerato superficialmente refuso da *modulazione*. L'osservazione del suo contesto d'uso ne rivela invece la natura tecnica e il significato:

Talvolta si uniscono in provetta seme ed ovuli dei donatori, si lascia iniziare in provetta la duplicazione dello zigote (morulazione) che successivamente si congela, per poi impiantare gli embrioni quando necessario.

A seguito di questa procedura si sono ottenuti 191 lemmi per il suffisso *-aggio* (a fronte di 1204 iniziali), 1906 lemmi per *-mento* (contro i 5053 iniziali), 1063 per *-tura* (da 3136) e 3009 con *-zione* (da 8905). La minor proporzione di lemmi considerati per *-aggio* può essere spiegata proprio per la maggior presenza di derivati denominali, i quali, come abbiamo precedentemente visto, non sono stati presi in considerazione in questo studio.

Dai dati così ricavati si è proceduto al calcolo della misura di produttività potenziale P e della sua variante tramite i modelli LNRE (specificatamente, attraverso il Zipf-Mandelbrot finito, fZM). Per le analisi si è utilizzato il software opensource R (R Development Core Team, 2003) e il pacchetto *zipfR* (Evert & Baroni 2007), nel quale sono già implementate le formule con i modelli parametrici.

Nella Tabella 1 si possono osservare i risultati relativi ad ogni suffisso con la misura tradizionale di produttività potenziale P . Come illustrato nella sezione 3, N indica il numero totale di *token* attestati con il suffisso in considerazione, V il numero di *type* distinti, V_1 il numero di *type* con frequenza 1, ovvero gli *hapax* di quel particolare suffisso. I dati sono stati ordinati in base alla misura P e, per facilitarne la leggibilità, i suoi valori sono stati moltiplicati per 100. Come si può osservare, il suffisso *-tura* risulta il più produttivo, mentre *-zione* si trova in ultima posizione. Un simile risultato va in direzione contraria rispetto a quanto osservato sia dagli studi precedenti, sia dal numero di *type* attestati nel *corpus*, che come si è visto nella sezione 3 è spesso utilizzato quale misura stessa.

Tabella 1 - *Misure tradizionali di produttività per i nomi d'azione dell'italiano contemporaneo*

		N	V	V_1	$P(*100)$
1	<i>-tura</i>	154980	1063	319	0,20583
2	<i>-aggio</i>	67236	191	48	0,07139
3	<i>-mento</i>	836449	1906	478	0,05715
4	<i>-zione</i>	1689009	3009	815	0,04825

Applicando il modello fZM, otterremo invece un ordine di produttività ben diverso, come si può osservare nella tabella 2. Qui la misura P è stata calcolata considerando come N quattro valori diversi, corrispondenti al numero di *token* attestati per ciascun suffisso. Quindi, al variare del valore considerato per N , l'ordine di produttività dei suffissi non cambia, sebbene siano osservabili differenze tra i valori di produttività.

Tabella 2 - *Misure parametriche di produttività per i nomi d'azione dell'italiano contemporaneo (moltiplicate per 100)*

	P	$E(V1)/N_{zione}$	$E(V1)/N_{mento}$	$E(V1)/N_{aggio}$	$E(V1)/N_{tura}$
1 -zione	0,04825	0,04825	0,07367	0,53649	0,27793
2 -mento	0,05715	0,03000	0,05715	0,41735	0,22091
3 -tura	0,20583	0,02712	0,05257	0,38085	0,20583
4 -aggio	0,07139	0,00030	0,00230	0,07139	0,02988

I risultati confermano la maggiore produttività del suffisso *-zione* osservata, seppure con metodi diversi, da tutti gli studi precedenti. Il discostarsi dei valori di P da tutti gli altri risultati ottenuti lascia presumere che questa non sia una misura appropriata per il rendimento. Le predizioni del modello fZM si sono dimostrate stabili a prescindere dalla grandezza del campione considerato e per questa ragione tale metodologia sembra essere la più affidabile.

5. Conclusioni

In questo lavoro si sono passate in rassegna le principali misure utilizzate per lo studio della produttività, mettendone in evidenza il rapido mutamento in pochi decenni. In particolar modo, si sono evidenziati i limiti della misura più utilizzata, la *produttività potenziale* (P , Baayen 1993) e si è mostrato come questi siano superabili mediante l'uso di modelli parametrici chiamati LNRE (Baayen 2001). Questi modelli permettono di comparare la produttività di affissi con frequenze molto diverse e, anche, tra più *corpora*. I risultati ottenuti sulla produttività dei nomi d'azione in italiano ne sono una prova.

A tal proposito, si è osservato che nell'italiano contemporaneo il suffisso *-zione* è il più produttivo tra i processi morfologici utilizzati nella formazione di nomi d'azione, seguito nell'ordine da *-mento*, *-tura* e in ultimo *-aggio*. Tuttavia, lo studio della loro produttività non si può ritenere completo. Bisognerebbe innanzitutto considerare anche i suffissi che qui (a causa del gran numero di dati analizzati) non sono stati considerati, ovvero, *-nza* e i derivati con suffisso zero. In secondo luogo, bisognerebbe comparare risultati tra diversi tipi di *corpora*. Come notato da Plag *et al.* (1999), i valori di produttività possono differire in base al genere testuale e al registro considerato.

Inoltre, questi studi dovrebbero andare di pari passo con l'osservazione dei domini di applicazione degli affissi, i quali possono fornirci una spiegazione quantomeno parziale della diversa produttività.

Per quanto riguarda le misure quantitative proposte, al di là della loro maggiore adeguatezza teorica e robustezza empirica, sarebbe auspicabile individuare ulteriori criteri per la loro validazione. Un'interessante prospettiva sarebbe offerta dal paragone con dati psicolinguistici. Diversi studi (ad es. Aronoff & Schvaneveldt 1978; Schröder & Mühleisen 2010) hanno stimato la produttività mediante giudizi di parlanti elicitati tramite diversi compiti, ma i risultati così ottenuti non sono stati confrontati direttamente con quelli ottenuti da *corpora*.

Riferimenti bibliografici

- Aronoff, Mark. 1976. *Word formation in generative grammar*. Cambridge, Mass.: MIT Press.
- Aronoff, Mark & Schvaneveldt, Peter. 1978. Testing Morphological Productivity. *Annals of the New York Academy of Sciences: Papers in Anthropology and Linguistics* 318. 106-114.
- Baayen, R. Harald. 1992. Quantitative aspects of morphological productivity. In Booij, Geert & Van Marle, Jaap (a cura di), *Yearbook of Morphology* 1991, 109-149. Dordrecht: Kluwer Academic Publishers.
- Baayen, R. Harald. 1993. On frequency, transparency, and productivity. In Booij, Geert & van Marle, Jaap (a cura di), *Yearbook of Morphology* 1992, 181-208. Dordrecht: Kluwer Academic Publishers.
- Baayen, R. Harald. 2001. *Word Frequency Distributions*. Dordrecht: Kluwer.
- Baayen, R. Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In Lüdeling, Anke & Kytö, Merja (a cura di), *Corpus Linguistics. An International Handbook*. Vol. 2, 899-919. Berlin: Mouton de Gruyter.
- Baayen, R. Harald & Renouf, Antoinette. 1996. Chronicling The Times: Productive lexical innovations in an English newspaper. *Language* 72. 69-96.
- Bauer, Laurie. 2001. *Morphological Productivity*. Cambridge: Cambridge University Press.

- Bauer, Laurie. 2005. Productivity: theories. In Štekauer, Pavol & Lieber, Rochelle (a cura di), *Handbook of word-formation*, 315-334. Dordrecht: Springer.
- Bazzanella, Carla. 2003. Nuove forme di comunicazione a distanza, restrizioni contestuali e segnali discorsivi. In Maraschio, Nicoletta & Poggi Salani, Teresa (a cura di), *Italia linguistica anno Mille - Italia linguistica anno Duemila. Atti del XXXIV Congresso della Società di Linguistica Italiana (SLI), Firenze, 19-21 ottobre 2000*, 403-415. Roma: Bulzoni.
- Berruto, Gaetano. 2005. Italiano parlato e comunicazione mediata dal computer. In Hölker, Klaus & Maaß, Christiane (a cura di), *Aspetti dell'italiano parlato*, 137-156. Münster: LIT Verlag.
- Carstairs-McCarthy, Andrew. 1992. *Current Morphology*. London: Routledge.
- Corbin, Danielle. 1987. *Morphologie dérivationnelle et structuration du lexique*, vol. 1. Tübingen: Niemeyer.
- Evert, Stefan. 2004. A simple LNRE model for random character sequences. In *Proceedings of the 7èmes Journées Internationales d'Analyse Statistique des Données Textuelles (JADT 2004), Louvain-la-Neuve, Belgium*, 411-422.
- Evert, Stefan & Baroni, Marco. 2007. *zipfR*: Word frequency distributions in R. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Posters and Demonstrations Session, Prague*, 29-32.
- Fiorentino, Giuliana. 2008. Action nouns and the infinitive in Italian. *Romanische Forschungen* 120(3). 3-28.
- Gaeta, Livio. 2004. Nomi d'azione. In Grossmann, Maria & Rainer, Franz (a cura di), *La formazione delle parole in italiano*, 314-351. Tübingen: Niemeyer.
- Gaeta, Livio & Ricca, Davide. 2002. Corpora testuali e produttività morfologica: i nomi d'azione italiani in due annate della Stampa (1996-1997). In Bauer, Roland & Goebel, Hans (a cura di), *Parallela IX. Testo-variazione-informatica/Text-Variation-Informatik. Atti del IX Incontro italoaustriaco dei linguisti, Salzburg, 1-4 novembre 2000*, 223-249. Wilhelmsfeld: Egert.
- Gaeta, Livio & Ricca, Davide. 2006. Productivity in Italian word formation: A variable-corpus approach. *Linguistics* 44(1). 57-89.
- Iacobini, Claudio & Thornton, Anna Maria. 1992. Tendenze nella formazione delle parole nell'italiano del ventesimo secolo. In Moretti, Bruno & Petrini, Dario & Bianconi, Sandro (a cura di), *Linee di tendenza nell'i-*

- taliano contemporaneo. Atti del XXV Congresso della Società di Linguistica Italiana (SLI), Lugano, 19-21 settembre 1991*, 25-55. Roma: Bulzoni.
- Lyding, Verena & Stemle, Egon & Borghetti, Claudia & Brunello, Marco & Castagnoli, Sara & Dell'Orletta, Felice & Dittmann, Henrik & Lenci, Alessandro & Pirrelli, Vito. 2014. The PAISÀ corpus of Italian web texts. In *Proceedings of the 9th Web as Corpus Workshop (WAC-9), in conjunction with EACL 2014, Gothenburg, Sweden*, 36-43.
- Plag, Ingo. 1999. *Morphological Productivity. Structural Constraints in English Derivation*. Berlin/New York: Mouton de Gruyter.
- Plag, Ingo. 2003. *Word-formation in English*. Cambridge: Cambridge University Press.
- Plag, Ingo. 2006. Productivity. In Aarts, Bas & McMahon, April M.S. (a cura di), *The Handbook of English Linguistics*, 537-556. Malden, MA: Blackwell.
- Plag, Ingo & Dalton-Puffer, Christiane & Baayen, R. Harald. 1999. Productivity and register. *English Language and Linguistics* 3(2). 209-228.
- Rainer, Franz. 2005. Constraints on productivity. In Štekauer, Pavol & Lieber, Rochelle (a cura di), *Handbook of word-formation*, 335-352. Dordrecht: Springer.
- Rossi, Fabio. 2010. Internet, lingua di. In *Enciclopedia dell'italiano*, Treccani online. ([http://www.treccani.it/enciclopedia/lingua-di-internet_\(Enciclopedia-dell%27Italiano\)/](http://www.treccani.it/enciclopedia/lingua-di-internet_(Enciclopedia-dell%27Italiano)/)) (Consultato il 5.1.2019).
- R Development Core Team. 2003. *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. <http://www.r-project.org/>.
- Schröder, Anne & Mühleisen, Susanne. 2010. New ways of investigating morphological productivity. *Arbeiten aus Anglistik und Amerikanistik* 35. 43-59.
- Štichauer, Pavel. 2009. Morphological productivity in diachrony: The case of the deverbal nouns in *-mento*, *-zione* and *-gione* in Old Italian from the 13th to the 16th century. In Montermini, Fabio & Boyé, Gilles & Tseng, Jesse (a cura di), *Selected Proceedings of the 6th Décembrettes*, 138-147. Somerville, MA: Cascadilla Proceedings Project.
- Thornton, Anna Maria. 1988. *Sui nomina actionis in italiano*, Pisa: Università di Pisa (Tesi di dottorato).
- Thornton, Anna Maria. 2005. *Morfologia*. Roma: Carocci.

