

Cenni di Morfologia II

Marco Baroni

29 maggio 2005

Email baroni@sslmit.unibo.it

Homepage del modulo <http://sslmit.unibo.it/~baroni/teaching.html>

Telefono 0453/374744

Ricevimento Martedì dalle 17:30 alle 19:30 presso lo Studio di Italiano e Linguistica (via Oberdan 2)

- In questo handout, prendiamo rapidamente in considerazione tre aspetti della morfologia di natura diversa da quelli trattati nello handout precedente.
- In particolare, fino ad ora abbiamo considerato le singole parole come entità autonome, mentre qui in 1 discuteremo di come una parola possa influenzare le caratteristiche morfologiche (e morfo-fonologiche) di parole relate attraverso l'effetto dell'*analogia*.
- In 2, proponiamo poi una classificazione di base delle lingue sulla base del tipo di processi morfologici che favoriscono.
- In fine, in 3 discutiamo (molto molto rapidamente) della morfologia nel contesto dell'*uso*, e cioè di cosa determina che certi processi siano utilizzati più di altri per formare parole nuove, della frequenza d'uso della parole, e di questioni relate.

1 L'analogia

- Nello handout precedente, abbiamo esaminato come parole complesse possano venire costruite attraverso processi concatenativi e non-contatenativi.
- Tuttavia, la forma di una parola complessa non è determinata solo dai morfemi che la compongono, ma anche dalla relazione che tale parola (e/o i morfemi che la compongono) ha con altre parole, che sono in qualche modo relate.

- In particolare, c'è una tendenza per la quale, sia nel corso della storia di una lingua, sia in un sistema sincronico, la forma di una parola/morfema/allomorfo viene influenzata dalla forma di parole/morfemi/allomorfi relati.
- Ci riferiamo a tale influenza con il termine di **analogia**.

1.1 Il livellamento analogico

- Il caso più tipico di analogia riguarda forme all'interno dello stesso paradigma, cioè forme che hanno il medesimo morfema lessicale, ma affissi distinti.
- In casi del genere, se il morfema lessicale viene rappresentato da allomorfi diversi in forme diverse del paradigma, ci può essere una tendenza ad estendere uno degli allomorfi a tutti i contesti, in modo da ridurre la variazione tra le parole all'interno del paradigma.
- Un esempio classico di questo fenomeno, noto come **livellamento analogico** (o livellamento paradigmatico) viene da paradigmi nominali latini quali quello di *honos* "onore".
- Per via di una restrizione fonotattica del latino arcaico, che vietava le /s/ intervocaliche, le forme flesse di questo paradigma sostituirono la /s/ finale con [r]: *honos, honoris, honori...*
- Dunque, per un certo periodo i paradigmi di questo genere avevano due allomorfi: l'allomorfo in [s] (*honos*) nel nominativo (cioè, come morfema lessicale libero), e l'allomorfo in [r] (*honor-*) nelle forme flesse (cioè, davanti ai suffissi flessivi).
- Tuttavia, alla lunga la spinta al livellamento paradigmatico ha fatto sì che l'allomorfo in [r] si estendesse anche al nominativo, portando al paradigma del latino classico, dove troviamo un solo allomorfo: *honor, honoris, honori...*
- Un altro esempio di livellamento paradigmatico viene da paradigmi verbali dell'yiddish, quale quello di *visn* "sapere", dove le forme del plurale avevano originariamente un allomorfo della radice diverso, poi sostituito con lo stesso allomorfo usato al singolare:¹
 - Prima del livellamento: *veys, veyst, veys, visn, vist, visn*.
 - Dopo il livellamento : *veys, veyst, veys, veysn, veyst, veysn*.

¹Esempi da [Albright 2002a].

1.2 Analogia tra morfemi lessicali distinti

- In altri casi, spinte analogiche vengono da parole o morfemi lessicali differenti ma relate/i (per es., verbi che appartengono ad una stessa “classe”).
- Per esempio, [Mioni 2001] osserva che errori quali *dassi* e *stassi* (invece che *dessi* e *stessi*) sono piuttosto comuni.
- Probabilmente, tali errori sorgono dall’analogia con le forme degli altri verbi in *-are*: *andare/andassi, cantare/cantassi, portare/portassi...*
- Un interessante caso di analogia di questo genere è stato scoperto da [Albright 2002b].
- Albright studia il seguente problema: data la prima persona di un verbo inventato, quale sarà la forma dell’infinito (e dunque la coniugazione) scelta da un parlante italiano?
- Se, per esempio, sentite la forma (*io*) *caffudo*, vi sembra più naturale che derivi da *caffudare, caffùdere, caffùdere* o *caffudere*?
- Ovviamente, in casi del genere c’è una preferenza per le forme in *-are*, visto che gran parte dei verbi italiani appartengono a questa classe (infatti, anche verbi importati di recente quali *formattare* o *resettare* tendono a finire nella classe in *-are*).
- Tuttavia, Albright mostra (attraverso un esperimento costruito con molta attenzione) che la scelta della classe dipende anche da spinte analogiche provenienti da verbi esistenti che assomigliano ai verbi inventati proposti.
- Così, per esempio, per la prima persona *aggiendo*, i parlanti trovano accettabile sia *aggiendare* che, cosa più interessante, *aggiéndere*.
- Questa intuizione è probabilmente dovuta al fatto che molti verbi la cui radice finisce in *-end-* appartengono appunto alla coniugazione in *-ere*: *accendere, apprendere, comprendere, prendere, rendere, stendere...*
- Anche per *caffudo* i parlanti studiati da Albright trovano accettabili sia *caffudare* che *caffùdere*, la seconda forma probabilmente modellata su forme come *preludere* e *colludere*.
- Invece, per *dolifico*, i parlanti trovano accettabile solo *dolificare* – chiaramente, un effetto della somiglianza con forme reali quali *pontificare, edificare, codificare...*²

²Si veda [Albright 2002b] per la lista completa degli stimoli usati, e per una discussione di come viene computata la similarità tra i verbi inventati e famiglie di verbi esistenti.

1.3 La distribuzione di /s/ al confine di morfema in italiano settentrionale: un caso di analogia?

- Come abbiamo visto in precedenza, il fonema /s/ in italiano settentrionale viene realizzato come [z] tra due vocali e davanti a consonante sonora, come [s] in tutti gli altri contesti.
- Questa distribuzione viene rispettata in maniera molto sistematica, anche nei prestiti (*[s]acher, O[z]aka, El Pa[z]o...*) e nelle parole inventate (*samo* e *satto* vengono lette con una [s], *pasa* e *laso* vengono lette con una [z]).³
- Tuttavia, c'è una classe di parole che costituiscono un'eccezione sistematica alla regola che /s/ intervocalica va realizzata come [z]: *pre[s]elezione, a[s]immetrico, a[s]ociale, pre[s]enile...* Che cosa hanno in comune queste parole?
- Non tutti i confini di morfema bloccano la sonorizzazione di /s/: *di[z]abile, ca[z]etta...*
- Cosa succede con la /s/ di *autobus, lapis, gas*, se aggiungiamo un morfema derivazionale che inizi per vocale?
- Una possibile interpretazione dei dati (proposta da [Kenstowicz 1997] e altri): /s/ viene prodotto come [s] *per analogia*, laddove esista una parola relata dove la stessa /s/ si trovi in inizio (o fine di parola).
- Dunque, *a[s]ociale* per analogia con *sociale, autobu[s:]ino* per analogia con *bus*, mentre *di[z]abile* o *ca[z]etta* non sono relate a parole dove le stesse /s/ vengano pronunciate come [s].
- Questo spiegherebbe anche *re[z]istere, pre[z]agio, de[z]erto*.
- Si notino però *ante[s]ignano, a[s]intoto, poli[s]emo, poli[s]indeto, pre[s]idio*, addirittura, per molti parlanti, *para[s]anghe*.
- E c'è variazione tra [s] e [z] in *coseno, resuscitare, risaltare* e molte altre parole la cui base esiste come parola autonoma.
- Probabilmente, le origini del fenomeno sono analogiche, ma nel sistema sincronico altri fattori (trasparenza semantica, lunghezza del prefisso) sembrano avere un peso maggiore (vedi [Baroni 2001]).⁴

³Vedi i dati riportati in [Baroni 2001].

⁴Una questione importante per la quale rimando alla recente, interessantissima tesi di dottorato di Adam Albright ([Albright 2002a]) è quella di come possiamo predire la *direzione* dell'analogia? Perché, per esempio, diciamo *a[s]ociale* e non *[z]ociale*?

2 Tipologia morfologica delle lingue

- Sapir (autore del famosissimo libro *Language*: [Sapir 1921]) ha proposto varie classificazioni delle lingue sulla base delle loro caratteristiche morfologiche.
- Presento qui una semplificazione della tipologia di Sapir, e sottolineo inoltre che si tratta più di tendenze che di differenze assolute (cioè, una lingua può essere più di un tipo o più di un altro, ma non esistono lingue che siano puramente isolanti, agglutinative o fusive.)⁵

2.1 Il tipo isolante

- Le lingue **isolanti** essenzialmente non hanno una morfologia: tutti i morfemi sono liberi, cioè non c'è distinzione tra morfema e parola.
- Il cinese mandarino tende fortemente a questo tipo.
- Si confrontino per es.:
 - wǒ mēn tān tǎin (io PLUR suonare piano) “Noi suoniamo il piano”.
 - wǒ mēn tān tǎin le (io PLUR suonare piano PASSATO) “Noi suonammo il piano”.
- Nelle lingue isolanti, per ovvie ragioni, l'ordine delle parole tende ad essere fisso.

2.2 Il tipo agglutinativo

- Nel tipo agglutinativo, a ciascun morfema tende ad essere associato un solo significato, e l'allomorfia è rara.
- Dunque, le lingue agglutinative (tra cui: ungherese, turco, finlandese, swahili, giapponese...) sono quelle dove le parole tendono più chiaramente ad avere una struttura a “trenino” – sono cioè caratterizzate da una lunga stringa di morfemi facilmente separabili.
- Per esempio, si considerino le seguenti forme del paradigma del verbo *soma* “leggere” in swahili:
 - ni-na-soma (io PRES leggere) “io leggo”.
 - u-na-soma (tu PRES leggere) “tu leggi”.
 - ni-li-soma (io PASS leggere) “io lessi”.
 - u-li-soma (tu PASS leggere) “tu leggesti”.
 - ni-ta-soma (io FUT leggere) “io leggerò”.
 - u-ta-soma (tu FUT leggere) “tu leggerai”.

⁵Baso la discussione che segue su [Jannedy et al. 1994], da cui prendo anche gli esempi.

2.3 Il tipo fusivo

- Purtroppo, l'italiano e le altre lingue indo-europee appartengono al tipo fusivo, che è caratterizzato dalla tendenza ad assegnare più di un significato per morfema, e da un alto grado di allomorfia.
- Da qui, i problemi a segmentare le parole in morfemi (come si segmentano *lessi* o *leggesti?*), e le difficoltà nell'apprendimento della morfologia di lingue straniere fusive.

3 Produttività e frequenza

- Non tutti i processi di derivazione hanno lo stesso grado di **produttività**, cioè, non tutti i processi hanno la stessa probabilità di venire usati per formare parole nuove.
- Per es., cfr. *ri-* vs. *re-*, *-are* vs. *-ire*, *-ità* vs. *-ezza*, *-zione* vs. *-nza*...
- Vari fattori determinano la maggiore o minore produttività di un affisso e del processo morfologico che ne fa uso.
- Un fattore importante è la *frequenza* con cui l'affisso viene usato in parole già esistenti.
- Intuitivamente, più sono le parole, e soprattutto le parole *trasparenti*, in cui un affisso viene usato, più sarà facile che il parlante decida di usare l'affisso per creare parole nuove.
- Si noti che quello che conta è il numero di parole diverse in cui un affisso capita, piuttosto che la frequenza di tali parole: cioè, un prefisso che capita in 100 parole poco frequenti sarà probabilmente più produttivo di un prefisso che capita in una o due parole frequentissime.
- In altre parole, ciò che conta è la **type frequency** (frequenza in una lista di parole), e non la **token frequency** (frequenza in un corpus).
- Per esempio, il fatto che la (o le) coniugazione(/i) in *-ere* dell'italiano contengano verbi molto frequenti come *essere*, *avere* e *potere* non basta a rendere tale(/i) classe(/i) produttiva(/e) – il fattore determinante che rende la coniugazione in *-are* la più produttiva è che ci sono un gran numero di verbi distinti (molti a bassissima frequenza) che appartengono a questa coniugazione (secondo [Albright 2002b], il 72% dei verbi italiani appartengono alla coniugazione in *-are*).
- Abbiamo tempo per un po' di speculazioni su token frequency, lessicalizzazione e opacità?
- Vedi, tra gli altri, [Bybee 1985], [Baayen e Lieber 1991], [Plag et al. 1999] e [Hay 2000] per una discussione più approfondita dei temi di produttività e frequenza.

3.1 La legge di Zipf

- Per finire, un'osservazione relata alla frequenza (intesa come token frequency) che chiunque si occupi di questioni linguistiche, dovrebbe sempre tenere presente: la cosiddetta **legge di Zipf**.
- In un paragrafo: In un corpus di dimensioni sufficientemente estese, il numero di parole che capitano una volta sola è approssimativamente pari alla metà delle parole nel corpus (formulazione presa da [Plag et al. 1999]).

Riferimenti bibliografici

- [Albright 2002a] A. Albright. 2002. *The identification of bases in morphological paradigms*. Tesi di Dottorato, UCLA.
- [Albright 2002b] A. Albright. 2002. Islands of reliability for regular morphology: Evidence from Italian. *Language*, in corso di pubblicazione.
- [Baayen e Lieber 1991] H. Baayen e R. Lieber. 1991. Productivity and English derivation: A corpus-based study. *Linguistics* 29: 801-843.
- [Baroni 2001] M. Baroni. 2001. The representation of prefixed forms in the Italian lexicon: Evidence from the distribution of intervocalic [s] and [z] in northern Italian. *Yearbook of Morphology* 1999: 121-152.
- [Bybee 1985] J. Bybee. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.
- [Hay 2000] J. Hay. 2000. *Causes and consequences of word structure*. Tesi di Dottorato, Northwestern University.
- [Jannedy et al. 1994] S. Jannedy, R. Poletto e T. Weldon (a cura di). 1994. *Language files*. Columbus: OSU Press.
- [Kenstowicz 1997] M. Kenstowicz. 1997. Base identity and uniform exponence: alternatives to cyclicity. In Durand e Laks (a cura di). *Current trends in phonology: Models and methods*. Salford: University of Salford. 363-394.
- [Mioni 2001] A. Mioni. 2001. *Elementi di morfologia generale*. Padova: Unipress.
- [Plag et al. 1999] I. Plag, C. Dalton-Puffer e H. Baayen. 1999. Morphological productivity across speech and writing. *English Language and Linguistics* 3: 209-228.
- [Sapir 1921] E. Sapir. 1921. *Language*. New York: Harcourt & Brace.