



## La sillaba come stabilizzatore di forze fonotattiche. Una modellizzazione\*

Basilio Calderone & Pier Marco Bertinetto  
Scuola Normale Superiore  
PISA  
*{b.calderone; bertinetto}@sns.it*

*(in corso di stampa sugli atti del 34° Congresso Internazionale di Studi  
della Società di Linguistica Italiana, Vercelli, settembre 2006)*

### 1. Sensibilità alla fonotassi

L'attitudine del parlante a registrare (più o meno implicitamente) gli aspetti statisticamente regolari e sistematici del linguaggio naturale è stata ampiamente dimostrata sul piano psicolinguistico (Cleeremans et al., 1998). Il presente lavoro si concentra su un aspetto delle regolarità fonotattiche: è stata studiata la stretta connessione tra la fonotassi, che regola la successione dei segmenti fonemati, e la struttura sillabica. Sfruttando tali regolarità, si è cercato di modellare computazionalmente il sistema sillabico in termini di attrazione/repulsione tra segmenti adiacenti. Cercheremo in primo luogo di chiarire come la sillaba (o meglio, il processo di sillabazione) sia motivata da costrizioni locali di natura fonotattica e come i due livelli (fonotattico e sillabico) interagiscano in maniera diretta, con conseguenze sullo sviluppo della competenza fonologica da parte del parlante. Sotto il profilo computazionale, il lavoro suggerisce strumenti per derivare l'organizzazione sillabica di una lingua a partire da informazioni sulla distribuzione dei fonemi.

Diversi approcci, propri del *machine learning*, sono stati avanzati per l'individuazione automatica dei confini sillabici: trasduttori a stati finiti (Kiraz and

---

\* I due autori hanno sviluppato di conserva questo progetto. Ai fini concorsuali, le responsabilità vanno così attribuite: BC è responsabile per i §§ 2-3; i §§ 1 e 4 sono da attribuirsi ad entrambi gli autori.

Möbius, 1998), *probabilistic context-free grammars* (PCFGs, Müller, 2002), alberi di decisione per classificazione (van den Bosch et al., 1998). Tali approcci, peraltro, trovano motivazioni meramente ingegneristiche (come implementazioni in sistemi *text-to-speech*) e sono di scarso interesse dal punto di vista (psico-)linguistico. La nostra ricerca poggia invece sulla nozione di ‘conoscenza fonotattica implicita’, già sperimentalmente dimostrata (Cassar & Treiman, 1997; Pacton et al., 2002). Tale conoscenza, intesa come capacità di generalizzazione e giudizio sulla *well-formedness* di sequenze fonematiche, emerge come prodotto collaterale dell’esposizione del parlante al dato linguistico e delle sue ricorrenze statistiche.

Ambizione di questo lavoro, tuttavia, non è solo di indagare come la fonotassi possa proiettarsi sul processo di sillabazione, ma anche di riprodurre in termini computazionali il graduale formarsi di questa competenza.

## 2. Metodologia e *Corpora*

In una prospettiva computazionale-simbolica, l’individuazione di possibili confini sillabici si riduce ad una semplice operazione di *parsing* ottenuta per mezzo di qualche algoritmo dedicato *rule-based*. In un approccio sub-simbolico, la sillabazione è ricondotta invece alla capacità di generalizzazione che il modello neurale, ad apprendimento ultimato, è in grado di esibire. In tale approccio, i principi che governano il processo di sillabazione e l’organizzazione della sillaba non sono definiti *a priori* dallo sperimentatore, ma appaiono come il risultato dell’attività simultanea di più unità elementari operanti in parallelo (nodi). La modellizzazione conferisce natura ‘emergentista’ alla sillaba, considerata come configurazione di attività emergenti derivate da effetti di tipo ‘collettivo’, effetti generati dalle interazioni tra le singole unità, mediante eccitamento/inibizione delle loro connessioni.

La nostra modellizzazione ha previsto l’utilizzo di una rete neurale *feedforward* addestrata tramite algoritmo di *back-propagation* (Rumelhart et al., 1986). A differenza di altri approcci che utilizzano sistemi connessionistici per l’analisi e la simulazione del comportamento sillabico (algoritmi generici su base neurale, Oudeyer 2001, sistemi dinamici, Laks 1995, reti ricorsive, Stoianov & Nerbonne 1998), il nostro modello, ad apprendimento supervisionato, va inteso come

un ‘filtro adattivo’ tra i contesti fonotattico e sillabico, capace quindi di ricondurre ogni segmento della catena fonematica alla struttura sillabica, espressa in termini di preferenze fonotattiche tra segmenti adiacenti.

Il sistema provvede dunque ad associare un *set* di patterns vettoriali, considerati come input al sistema (contesto fonotattico), con un *set* di patterns di output (contesto sillabico; cf. Figura 1). Ogni segmento via via considerato in Focus è codificato tenendo conto del contesto fonotattico, sia a sinistra sia a destra. Tale accorgimento fa sì che, in fase di *training*, ogni segmento sia elaborato in corrispondenza del suo intorno fonotattico entro la parola. Il sistema neurale è in grado di ricondurre ogni segmento, considerato in un contesto fonotattico a finestra costante (coi quattro segmenti a sinistra e a destra<sup>1</sup>), ad un contesto sillabico a finestra variabile, in cui il segmento viene restituito assieme ai segmenti costituenti l’unità sillabica.

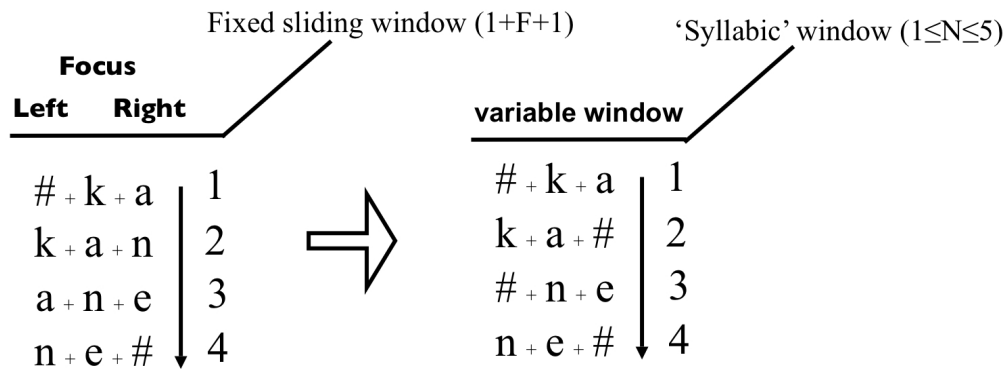


Figura 1. Contesto fonotattico ‘a finestra fissa’ (qui, per semplicità, ridotto ad una sola posizione a sinistra/destra del segmento in Focus) e contesto sillabico ‘a finestra variabile’ (per l’italiano, max. cinque segmenti). La parola qui considerata è *cane*. Ad un input di segmenti fonematici considerati nel loro contesto fonotattico, il sistema fornisce in output un contesto sillabico.

Sono stati confezionati tre diversi *corpora* di addestramento per italiano, spagnolo ed inglese. Ogni corpus è stato concepito per rappresentare in maniera molto

<sup>1</sup> In fase preliminare, sono state condotte diverse simulazioni, considerando una ‘finestra fonotattica’ allargata, con un numero di posizioni, a sinistra/destra del segmento in Focus, maggiori di quattro. Nel caso dello spagnolo, la dimensione della ‘finestra fonotattica’ è stata ridotta a tre posizioni-segmenti a sinistra/destra del Focus.

parsimoniosa i vari tipi sillabici della lingua. Il corpus italiano è formato da 83 parole per 51 tipi sillabici, quello spagnolo da 160 parole per 36 tipi sillabici, quello inglese da 159 parole per 78 tipi sillabici (Tabella 1).

Ogni segmento di input è stato codificato utilizzando una rappresentazione distribuita di tipo binario. Sette classi naturali sono state adottate per specificare la natura di ogni segmento: **V** (vocale), **G** (approssimante), **L** (liquida), **N** (nasale), **F** (fricativa), **O** (occlusiva) e **A** (rafforzata: raccoglie affricate dentali e foni palatali rafforzati).<sup>2</sup>

L'input per l'addestramento del sistema è stato creato nella maniera seguente: l'*i*-esimo segmento di una parola del *corpus* è rappresentato da un vettore binario  $\mathbf{V}_i$  a 7 componenti, che definisce la classe di appartenenza. Pertanto una parola composta da *n* segmenti sarà rappresentata dagli *n* vettori binari  $\mathbf{v}_1, \dots, \mathbf{v}_n$ . Il vettore di input  $\mathbf{V}_i$ , che denomineremo conteso fonotattico di Focus  $\mathbf{v}_i$ , è dato da:

$$\mathbf{V}_i = (\mathbf{v}_{i-4} \dots \mathbf{v}_{i-1} \mathbf{v}_i \mathbf{v}_{i+1} \dots \mathbf{v}_{i+4})$$

ove agli indici minori di 1 e maggiori di *n* corrisponde un vettore nullo. Data la natura supervisionata del sistema, un vettore binario  $\mathbf{W}_i$  dello stesso ordine di componenti è stato creato come risposta di output per ogni vettore di contesto fonotattico  $\mathbf{V}_i$ . Il vettore  $\mathbf{W}_i$  fornisce l'informazione sulla natura sillabica del segmento Focus  $\mathbf{v}_i$  per la parola in questione, marcando la presenza o meno di altri segmenti adiacenti, responsabili della struttura sillabica assieme al Focus.

	Parole	Tipi sillabici
Italiano	83	51
Inglese	159	78
Spagnolo	160	36

Tabella 1. Parole e tipi sillabici all'interno dei *corpora*.

## 2.1 Addestramento e Test

<sup>2</sup> Due ulteriori classi sono state aggiunte *ad hoc* per i dati inglese: W (vocale rilassata) e R (liquida sillabica).

Diverse architetture sono state progettate per verificare la minimizzazione dell'errore ad apprendimento ultimato. Sono state testate reti *feedforward* con 7, 10 e 15 neuroni entro lo stato nascosto (*hidden layer*) e reti a due livelli, senza strato nascosto. Quest'ultima architettura è sembrata la più promettente in termini di rendimento per le nostre finalità (vedi Figura 2).

Nella fase di addestramento, i coefficienti di connessione del sistema neurale sono inizializzati in maniera *pseudo-random* e gradualmente modificati, per minimizzare l'errore, mediante l'algoritmo *back-propagation*. Ad apprendimento ultimato (25000 cicli e tasso di apprendimento 0.4), la percentuale di sillabazioni corrette, per le forme presenti nei *corpora*, è stata del 99.2%.

Nella nostra modellizzazione, l'unità sillabica è concepita come un attrattore di forze fonotattiche, caratterizzato da valori di repulsione/attrazione con i diversi segmenti fonologici. I diversi tipi sillabici sono attratti dai vari contesti fonotattici (leciti) con intensità variabile. Una struttura sillabica è quindi definita dalle mutue attrazioni che i vari segmenti intrattengono al suo interno. L'accettabilità o meno di una sillaba è decisa dall'equilibrio di attrazione che i vari segmenti stabiliscono gli uni con gli altri entro il contesto locale.

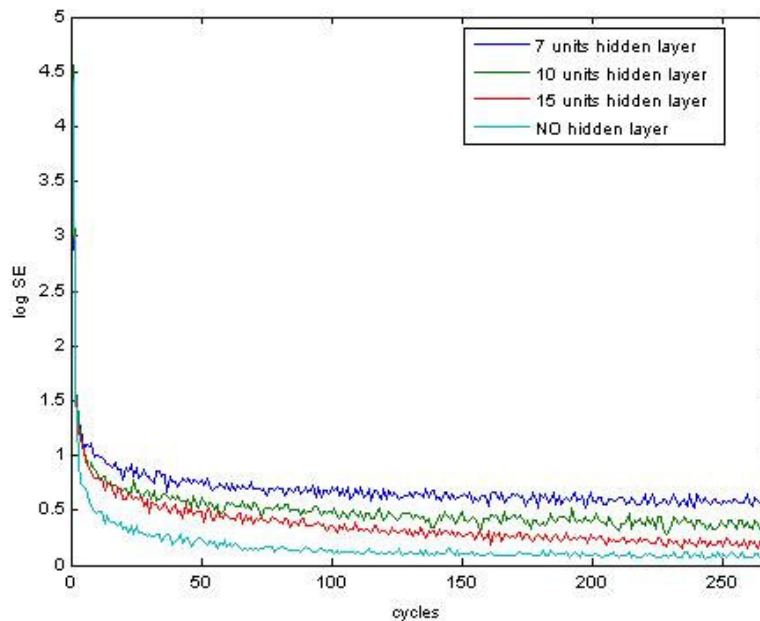


Figura 2. Errore quadratico medio (*Mean Square Error*) in scala logaritmica, calcolato per diverse tipologie di architetture ad apprendimento concluso (corpus italiano).

### 3. Risultati

#### 3.1 Generalizzazione

Il nostro modello definisce i valori di uscita all'interno di un *continuum* numerico reale tra 0 e 1: un'uscita di sillabazione non sarà mai pienamente 0 o 1, ma loro approssimazioni (ad es., 0.012 e 0.987). Ciò consente analisi quantitative raffinate all'interno dei possibili confini sillabici.

Un esempio visivo di sillabazione è dato dalla Figura 3 per la parola italiana /sillaba/ e quella inglese /bifor/. All'interno della parola, ogni segmento esibisce una propria preferenza fonotattica nella scelta dei segmenti per la costruzione di unità sillabiche. La sillaba emerge dunque come uno 'stato di equilibrio' tra le forze fonotattiche mutuamente esercitate dai segmenti. Da tale 'gioco di attrazioni', la sillaba affiora come unità di un livello di rappresentazione ulteriore (rispetto a quello fonotattico), su cui vengono proiettate e organizzate le sequenze fonemiche. Anziché una guida-supervisione, atta a definire l'accettabilità delle sequenze di segmenti, la sillaba è insomma il mero prodotto superficiale delle restrizioni fonotattiche profonde.

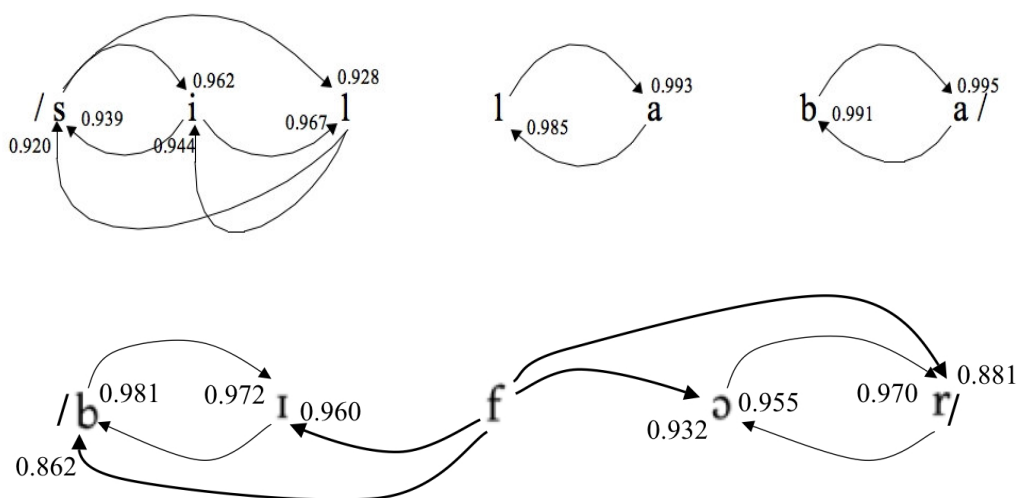


Figura 3. Attrazioni fonotattiche per le parole /sillaba/ e /bifor/. Si noti l'ambisillabicità di /f/ per la parola inglese.

Sulla base della conoscenza acquisita, il sistema è stato indotto a compiere due diversi processi di generalizzazione:

- a) adattare la conoscenza fonotattica a dati simili ma non identici; ossia, riconoscere tipi sillabici già appresi (presenti nei *corpora*), considerandoli tuttavia in contesti fonotattici differenti da quelli dell'apprendimento;
- b) adattare la conoscenza a dati inediti; ossia, riconoscere nuovi tipi sillabici, non presenti nell'addestramento.

Nel primo caso, il sistema ha dimostrato una robusta capacità di generalizzazione, riuscendo a riconoscere le 'vecchie' unità sillabiche nei nuovi contesti fonologici. Per es., il costrutto sillabico FGV, presente nel *corpus* italiano solo nell'item /swoi/ (FGV·V), è stato correttamente individuato in un contesto fonemico totalmente diverso come /graffjo/ (OLVF·FGV), sulla base delle preferenze fonotattiche espresse dai singoli segmenti per la costruzione della sillaba.

Nel secondo caso, nuovi tipi sillabici sono stati scoperti dal sistema per tutti i *corpora* studiati (italiano, spagnolo, inglese). Testato su forme inedite, il sistema ha mostrato capacità di generalizzazione riuscendo, sulla base della conoscenza appresa, a produrre una plausibile sillabazione. Ne è un esempio il costrutto FVG, individuato come unità sillabica in parole italiane come /fewdo/ (FVG·OV) o /sawro/ (FVG·LV) presentate per la prima volta alla rete. L'abilità di scoprire nuovi tipi sillabici è stata confermata anche in riferimento agli altri due *corpora*, spagnolo e inglese, di cui per esigenze di sintesi non riportiamo qui i risultati.

Il sistema sembra quindi possedere una strutturata e coerente conoscenza di quella che potrebbe essere la fonotassi di italiano, spagnolo e inglese, riuscendo a scomporre nuove ed inedite sequenze fonematiche in opportune costellazioni di attrazione (sillabe). In particolare, i valori di attrazione/repulsione che ogni segmento esprime nei confronti dei segmenti adiacenti, e che stanno alla base dell'accettabilità delle unità sillabiche, consentono di indagare quantitativamente contesti fonotattici critici, la cui scansione non appare affatto scontata.

### 3.2 Lo status di /s/+ C

É stato in particolare indagato il comportamento del nesso consonantico /s/+C all'interno di parola. Tale nesso non compariva entro le parole dei tre *corpora* di

addestramento, ma è stato volutamente introdotto come forma inedita per testare il rendimento del sistema.

Di contro all'ipotesi corrente circa l'eterosillabicità del nesso /s/+C in italiano (Marotta, 1995), è stata avanzata l'idea dell'indecidibilità di tale scansione. L'ipotesi, elaborata sul piano sia teorico sia sperimentale (Bertinetto, 1996 e 2004), suggerisce che la scansione di /s/+C sia governata, in italiano, da principi non deterministici: in pratica, i parlanti oscillano fra tautosillabicità ed eterosillabicità. Affidare il nesso /s/+C al sistema costituisce pertanto un valido test per valutarne le doti di generalizzazione.

Il sistema ha fornito chiari indizi di eterosillabicità per /s/+C in spagnolo e in inglese, in linea con le attese teoriche. Una diversa classificazione è invece emersa per l'italiano. La rete neurale ha collocato il nesso a metà strada tra eterosillabicità e tautosillabicità, in una zona di equilibrio caratterizzata da valori sfumati (ad es., 0.543 o 0.496). Nel caso di /pasta/ (Figura 4), il sistema ha evidenziato le multiple attrazioni fonotattiche di /s/ verso le due sillabe adiacenti, con valori 'opachi' rispetto all'alternativa attrazione/repulsione. Possiamo dunque concludere che, nella nostra modellazione, il nesso /s/+C non risulta propriamente né eterosillabico né tautosillabico.

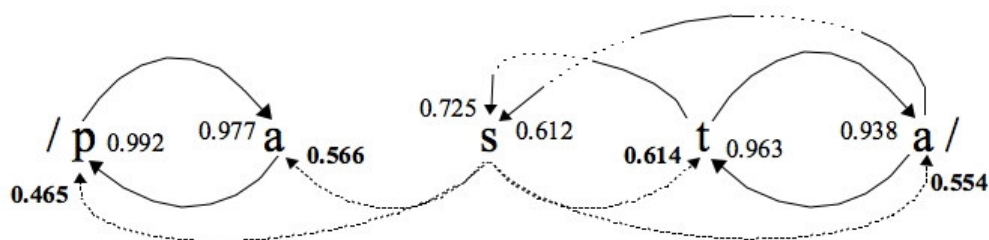


Figura 4. Generalizzazione del nesso /s/+C in italiano. Le connessioni 'a punti' rappresentano le attrazioni di /s/ verso i segmenti adiacenti.

### 3.3 Comportamento sillabico

Ad apprendimento ultimato, una nuova serie di generalizzazioni è stata condotta, al fine di registrare e analizzare il possibile 'comportamento sillabico' dei singoli segmenti. Sono stati creati sette vettori 'Focus'  $V_i$ , rappresentanti ognuno la propria classe ( $i = F, O, \dots, A$ ). Dal processo di generalizzazione è stato quindi



prodotto, per ognuna delle sette categorie, un nuovo vettore  $W_i$ , che definisce le preferenze sillabiche della categoria in Focus rispetto al suo contesto sinistro/destro.

I dendrogrammi delle Figure 5 e 6 illustrano i rapporti tra le sette categorie, rispettivamente per il contesto sinistro/destro. La metrica adottata per evidenziare le similarità è quella per cui la distanza fra due vettori  $W'$  e  $W''$  è data dall'angolo compreso tra i rispettivi versori (*cosine distance function*).

Si noti come il sistema differenzi nettamente, per entrambi i contesti, il comportamento sillabico delle vocali da tutti i segmenti prettamente consonantici. È interessante constatare come la classificazione dei segmenti del contesto sinistro appaia governata dal principio di sonorità, che raggruppa vocali e approssimanti (V e G), separandole dalle restanti categorie sulla base della forza consonantica. Meno felice appare invece la classificazione ottenuta per il contesto destro.

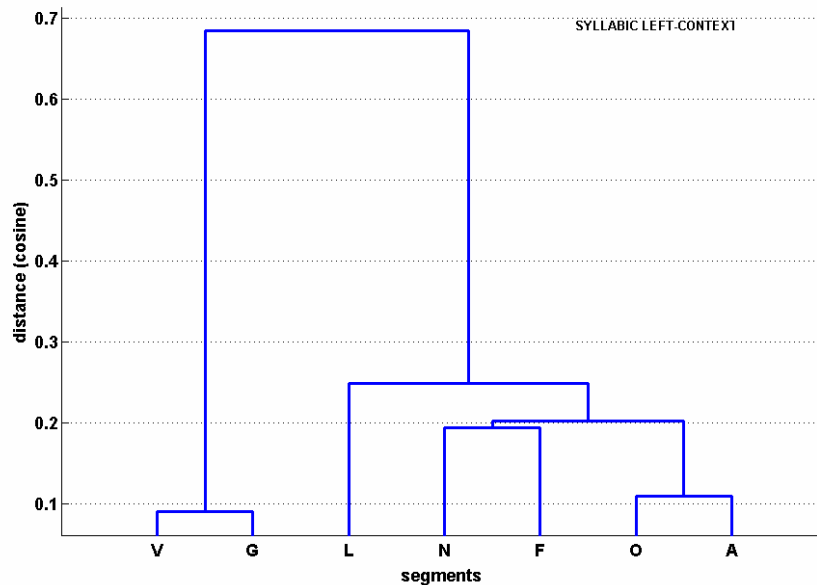


Figura 5. Similarità di ‘comportamento sillabico’, per il contesto sinistro, di ognuna delle sette categorie naturali.

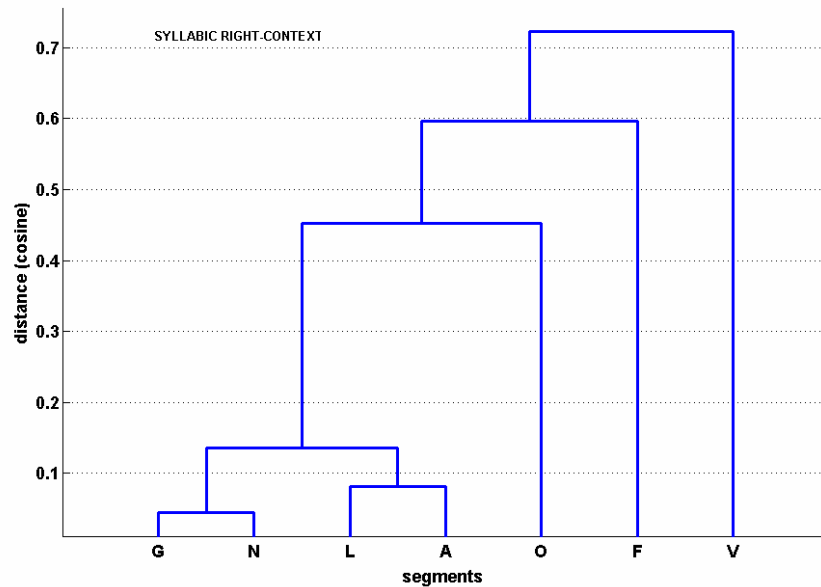


Figura 6. Similarità di ‘comportamento sillabico’, per il contesto destro, di ognuna delle sette categorie naturali.

## 4. Conclusioni

La nostra modellizzazione ha permesso di esprimere computazionalmente le preferenze fonotattiche che i vari segmenti esibiscono mutuamente per la realizzazione delle unità sillabiche. Il nostro sistema può dunque operare anche da sillabatore automatico, riuscendo a riconoscere e marcare i confini di sillaba all’interno di sequenze fonematiche. La possibilità di indagare il *continuum* entro cui si realizza la sillabazione permette inoltre, tramite processi di generalizzazione, di compiere analisi raffinate sui casi in cui il confine di sillaba risulta incerto.

In una prospettiva cognitiva, il lavoro di modellizzazione ha cercato di porre in luce la graduale formazione della competenza fonotattica implicita, che il parlante sviluppa in rapporto ai dati linguistici cui è esposto. Tale competenza, di natura non esplicita né dichiarativa, è stata intesa, nel nostro caso, come co-prodotto di sistematiche regolarità statistiche e di proiezioni fonotattiche (locali) su rappresentazioni sillabiche. Proiettare la fonotassi dell’italiano, dello spagnolo e dell’inglese sul livello sillabico proprio di ciascuna lingua ha permesso di definire le

costrizioni e le ‘liceità’ fonotattiche, per le quali una data sequenza fonematica risulta sillabicamente interpretabile.

Un perfezionamento del sistema, ancora in fase di sviluppo, consisterà nell’adottare una rappresentazione del segmento fonologico che prescindendo dalle classi naturali (vocale, nasale, occlusiva, ecc), e faccia invece riferimento o, direttamente, alla forza consonantica espressa da ogni segmento fonematico, oppure ai componenti minimi (tratti distintivi) che ne specificano la natura fonologica.

## 5. Bibliografia

- Cassar, Marie & Rebecca Treiman. 1997. The beginnings of orthographic knowledge: children’s knowledge of double letters in words. *Journal of Educational Psychology*, 89: 631- 644.
- Bertinetto, Pier Marco. 1996. Psycholinguistic evidence for syllable geometry: Italian and beyond. In Rennison & Kähnhammer (eds.), *Phonologica 1996. Syllables!?* Holland Academic Graphics:1-28.
- Bertinetto, Pier Marco. 2004. On the undecidable syllabification of /sC/ clusters in Italian: Converging experimental evidence. *Italian Journal of Linguistics / Rivista di Linguistica*, 16:349-372.
- Cleeremans, Axel, Arnaud Destrebecqz & Maud Boyer. 1998. Implicit learning: news from the front. *Trends in Cognitive Sciences*, 2: 406-416.
- Goldsmith, John. 1992. Local Modeling in Phonology. In Davis (ed.), *Connectionism: Theory and Practice*. Oxford, OUP:229-246.
- Joanisse, Marc. 1999. Exploring syllable structure in connectionist networks. In *Proceedings of the XIVth International Congress of Phonetic Sciences*, San Francisco, CA.

- Kiraz, George Anton & Bernd Möbius. 1998. Multilingual syllabification using weighted finite-state transducers. In *Proceedings of the Third International Workshop on Speech Synthesis*, Jenolan Caves, Australia, 71-76.
- Laks, Bernard. 1995. A connectionist account of French syllabification. *Lingua*, 95:349-372.
- Marotta, Giovanna. 1995. La sibilante preconsonantica in italiano: Questioni teoriche ed analisi sperimentale. In Ajello & Sani (eds.), *Scritti linguistici e filologici in onore di Tristano Bolelli*. Pisa, Pacini:393-437.
- Müller, Karin. 2002. Probabilistic context-free grammars for phonology. In *Proceedings of the ACL-02 workshop on Morphological and phonological learning - Volume 6*, Morristown, NJ: Association for Computational Linguistics, 70-80.
- Oudeyer, Pierre-Yves. 2001. Origins and Learnability of Syllable Systems: a Cultural Evolutionary Model. In Collet, Fonlupt, Hao, Lutton, Schoenauer (eds.), *Artificial Evolution*, London: Springer Verlag: 143-155.
- Pacton, Sébastien, Michel Fayol, Pierre Perruchet. 2002. The acquisition of untaught orthographic regularities in French. In Vervhoeven, Elbro & Reitsma (eds.), *Precursors of functional literacy*. Amsterdam: John Benjamins.
- Rumelhart, David. E., James McClelland & the PDP Research Group. 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Volumes 1 and 2. Cambridge, MA: MIT Press.
- Stoianov, Ivelin & John Nerbonne. 1998. Modeling the phonotactics of Natural Language Words with SRNs - part two: Exploring phonetic data representations. In *CLIN'98: Computational Linguistics in Netherlands*. Leuven, Belgium.

- Treiman, Rebecca, Brett Kessler, Stephanie Knewasser, Ruth Tincoff, Michael Bowman. 2000. English speakers' sensitivity to phonotactic patterns. In Broe and Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*. Cambridge, England: Cambridge University Press.
- van den Bosch, Antal, Ton Weijters, Walter Daelemans. 1998. Modularity in inductively-learned word pronunciation systems. In *Proceedings of the Third Conference on New Methods in Language Processing (NeMLaP3), and the Second Conference on Natural Language Learning (CoNLL98)*, Sydney, Australia, 185-194.