



UNIVERSITÀ DI PISA

**DIPARTIMENTO DI FILOLOGIA, LETTERATURA E LINGUISTICA
CORSO DI LAUREA MAGISTRALE IN INFORMATICA UMANISTICA**

TESI DI LAUREA MAGISTRALE

**Un approccio computazionale alla complessità
della derivazione: il caso dei derivati italiani in
*-zione, -tore, -ico***

**Candidata:
Flavia Basso**

**Relatore:
dr. Claudia Marzi**

Anno Accademico 2020/2021

Indice

1	INTRODUZIONE	5
2	INQUADRAMENTO TEORICO	11
2.1	Una nuova scienza del linguaggio	12
2.2	I modelli simbolici della teoria rappresentazionale	14
2.2.1	L'approccio duale	20
2.3	La rappresentazione dell'informazione	21
2.3.1	L'informazione morfologica	23
2.4	Evidenza comportamentale	26
2.5	Modelli computazionali del lessico mentale	27
2.5.1	La plausibilità biologica: neuroni e reti neurali artificiali	30
2.5.2	Apprendimento non-supervisionato: mappe auto-organizzanti	33
2.6	Problemi teorici della derivazione	35
2.6.1	Effetti di frequenza	37
2.6.2	Numerosità e produttività	38
3	METODI E DATI	41
3.1	Risorse e strumenti: i corpora	42
3.1.1	Il DerIvaTario e il CoLFIS	42
3.2	Il dataset di addestramento: i criteri di selezione	44
3.2.1	Criteri di annotazione	45
3.3	La Temporal Self-Organising Map	45
3.3.1	Architettura della rete	47

3.3.2	Algoritmo di apprendimento	48
3.3.3	Accuratezza: <i>recoding</i> e <i>recall</i>	51
3.3.4	Predizione	52
4	ANALISI DEI RISULTATI	53
4.1	Le dinamiche di <i>learning</i> e <i>processing</i>	54
4.2	Il ruolo della frequenza dell'input	62
4.3	Il ruolo della famiglia	73
5	DISCUSSIONE E CONSIDERAZIONI FINALI	83
5.1	Percezione della struttura morfologica	83
5.2	Una riflessione teorica	85
5.3	Possibili sviluppi	86
	BIBLIOGRAFIA	86
	APPENDICE	95

Capitolo 1

INTRODUZIONE

*Dum taxat rerum magnarum parva potest
res exemplare dare et vestigia notitiae*

Lucrezio, De rerum natura – Libro II

Tra le specificità della natura umana, quella del linguaggio costringe gli studiosi a cogliere la grande sfida contemporanea, ovvero di comprendere in che rapporto il linguaggio stesso è col pensiero. Un cospicuo moltiplicarsi di studi su come sia rappresentato, compreso ed elaborato il linguaggio, ha mostrato negli ultimi cinquant'anni che la complessità della mente va ben oltre quanto si sia immaginato in passato, toccando aree di ricerca distanti, eppure complementari nello spiegare i processi mentali in relazione alle strutture della lingua. Se è vero che il linguaggio rappresenta un punto di accesso *privilegiato* allo studio delle strutture mentali (Rizzi 1994), scavalcata la storica convinzione che *tutto* ciò che lo caratterizza sia arbitrario e culturale, il mondo scientifico si è addentrato in una controversia interessante che – nell'avvicinarsi di teorie – può risolversi in due principali linee di pensiero. Da un lato la convinzione che mente e linguaggio siano facoltà uniche e soprattutto innate degli esseri umani, dall'altro l'idea che la mente sia da considerare come frutto dell'apprendimento in base all'attivazione di diversi sistemi posti in rete. Con particolare riguardo a questa intuizione, il tradizionale approccio generativista viene superato da modelli computazionali nuovi, proposti come preziose miniature dei processi mentali coinvolti nei compiti linguistici.

La grammatica generativo-trasformativa, infatti, non risulta più essere un modello adatto ad analizzare i processi di produzione e di percezione del linguaggio, in quanto lascerebbe inspiegate numerose variabili di tipo costruttivo e funzionale circa il coinvolgimento della memoria. Queste problematiche sono invece parzialmente risolte dall'approccio connessionista, il cui primo modello di Rumelhart e McClelland (1986, 1987) costituisce una ispirazione importante per le simulazioni del comportamento osservabile, seppur nella estrema semplificazione del loro modello. Sono le recenti evidenze neurofisiologiche a porre enfasi sul substrato neurofisiologico che supportano una visione multi-livello e integrata della competenza lessicale del parlante (si veda ad esempio Catani, Jones e Ffytche 2005), ispirando i modelli psico-computazionali dell'elaborazione del dato linguistico (Elman 2004; Westermann e Plunkett 2007, tra altri). Se in termini *gestaltisti* un fenomeno è da considerarsi sovraordinato rispetto alla somma dei suoi componenti, entrando nel merito degli studi sul linguaggio è evidente l'impossibilità di ridurre tutte le manifestazioni linguistiche, e per tutti i livelli di analisi, ad un solo modello che possa spiegarne l'intero funzionamento; risulta quindi necessario specializzare il campo di osservazione fino ad arrivare ad un quesito sufficientemente semplice da permettere di essere indagato sperimentalmente. Una prima necessaria restrizione di campo è stata dunque quella di mirare ad un solo sottodominio del fenomeno linguistico e di chiedersi se quel dominio possa avere una rappresentazione autonoma al livello dell'attività neuronale. Nello specifico del presente lavoro, l'elaborazione dell'input linguistico e alcuni principi di base dei sistemi neuro-fisiologici soggiacenti ai suoi meccanismi, pur essendo complessi sul piano teorico e sperimentale, sono indagati in un approccio interdisciplinare avvalendosi degli strumenti della psicolinguistica computazionale¹ per verificare e discutere un modello che simuli in particolar modo la percezione della struttura morfologica di carattere più strettamente DERIVAZIONALE. Nel caso specifico di questa tesi i corpora presi in considerazione sono il *derIvaTario* e il *Colfis* (Talamo, Celata e P. Bertinetto 2016).

¹La psicolinguistica computazionale ha l'obiettivo di produrre modelli esplicativi mediante simulazioni computazionali, in grado di replicare alcuni dei processi cognitivi osservabili nei parlanti durante l'esecuzione di compiti linguistici. Dal punto di vista delle tecniche computazionali, adotta principi e meccanismi cognitivamente ispirati, riconducibili, ad esempio, alle dinamiche di plasticità hebbiana.

I principali interrogativi posti nello sviluppo di questa tesi vertono sulla percezione della struttura morfologica di forme derivate dei suffissi in *-zione*, *-tore*, *-ico*. Questi suffissi, con le relative allomorfie, presentano una diversa consistenza funzionale tale da poter influenzarne la percezione da parte del parlante, che si tenterà di modellare in una dinamica dipendente dall'elaborazione stessa. Si tenterà, pertanto, di comprendere se vi sia una diversa percezione della struttura interna delle parole derivate, sia in base alla diversa distribuzione dell'input, sia in base ad una eventuale diversa salienza dei suffissi derivazionali in esame.

Sebbene sembri ancora vivo il dibattito sull'elaborazione morfologica tra modelli simbolici e modelli sub-simbolici di rappresentazione (Pinker e Ullman 2002), è possibile avere un riscontro positivo dei tentativi di implementazione di alcuni principi di base dei sistemi neurali con architetture neurali ricorrenti per modellare quantitativamente le dinamiche di apprendimento e elaborazione lessicali variando le condizioni di input, quali, ad esempio, le *Temporal Self-Organising Maps* (o TSOM) (Ferro, Marzi e Pirrelli 2011)².

La proposta prende avvio da alcune delle posizioni più influenti degli ultimi anni ed è formulata volgendo l'attenzione ai dati della lingua italiana, tenendo presente contributi relativamente recenti (Pirrelli e Battista 2000). Sulla base di un solido inquadramento teorico l'idea è di gettare la base sperimentale dalla quale partire per replicare alcune evidenze e da queste proporre un modello esplicativo. L'interesse è non solo di capire quali fattori nella formazione delle parole possono influenzare la percezione del parlante di un certo suffisso, ma anche di determinare se c'è un legame tra la complessità morfologica e gli altri elementi linguistici che la caratterizza, come ad esempio l'effetto di frequenza – sia *token* che *type frequency*.

Il Capitolo 2 della presente tesi si apre con inquadramento teorico della disciplina di riferimento e dei diversi modelli di rappresentazione, passando in rassegna i domini empirici della ricerca psicolinguistica: il lessico mentale, l'acquisizione della competenza morfologica e il ruolo che ha nella percezione e nella produzione lessicale. Sono indagate soprattutto le conseguenze empiriche e teoriche di un modello integrato e distribuito del lessico menta-

²Questo tipo di rete neurale artificiale simula i processi di apprendimento per auto-organizzazione, estraendo informazione dall'input classificandone i pattern ricorrenti. Caratterizzata da un apprendimento non supervisionato, la rete riceve come input parole rappresentate da serie temporali di simboli somministrati uno alla volta.

le, passando dall'analisi psicolinguistica del fenomeno a una sua descrizione algoritmica di cui le simulazioni computazionali risultano preziose per la definizione di una relazione tra il comportamento dei parlanti e principi generali del fenomeno indagato.

In particolare, è fatto riferimento al paradigma connessionista e agli elementi innovativi ad esso legati, con descrizione della struttura dei modelli neurali, in cui sono mostrati il loro funzionamento e il loro significato in termini matematici e statistici³. In questi termini un approccio psico-computazionale arricchisce le scienze cognitive classiche, rappresentando la ripresa contemporanea di questa impostazione radicalmente modellistica, ponendosi come ponte tra livelli della gerarchia di Marr (1979).

Il Capitolo 3 mira a enucleare tutta la fase sperimentale, introducendo le risorse e gli strumenti utilizzati per la creazione del dataset di addestramento, delucidando i criteri di selezione dei dati estratti dal corpus del *derIvaTario* dell'Italiano e il tipo di esperimenti implementati con la rete neurale TSOM, della quale saranno dati alcuni dettagli del funzionamento. Gli esperimenti saranno condotti al fine di determinare se i suffissi e *pattern* morfologici svolgono un ruolo nell'accesso e nell'elaborazione di parole complesse italiane e, più precisamente, se la salienza del suffisso e la grandezza della serie e della famiglia influenzano l'elaborazione. Saranno sottoposti alla rete parole morfologicamente complesse per i tre suffissi di diversa salienza *-zione*, *-tore*, *-ico* (comprensivi di allomorfi), le cui serie presentano un diverso grado di consistenza funzionale, al fine di confrontare l'effetto della grandezza della famiglia con quello della grandezza della serie, ovvero quando una base si combina con diversi suffissi derivazionali e quando un suffisso si combina con più basi. Un primo esperimento sarà effettuato tramite un addestramento con frequenze uniformi, al fine di individuare gli andamenti della TSOM relativamente alle epoche di addestramento della rete (per quante volte viene reiterata la somministrazione dell'input) per un numero fisso di occorrenze delle parole (il numero di volte in cui ogni forma di input viene sottoposta alla rete), determinandone le percentuali di accuratezza per i compiti di ricodifica e recupero (*recall*). Un secondo addestramento con frequenze corpus-based servirà invece a definire il ruolo della frequenza dell'input nella modu-

³La pubblicazione manifesto di *Parallel Distributed Processing* (James L McClelland, David E Rumelhart, Group et al. 1986) è un tale momento di svolta nello studio della mente, da introdurre un paradigma innovativo all'interno della scienza cognitiva, il connessionismo.

lazione degli effetti risultanti dal primo esperimento, mentre un terzo esperimento indagherà specificamente il ruolo delle famiglie, ovvero delle relazioni tra forme riconducibili alla stessa base.

Il Capitolo 4 è dedicato all'analisi quantitativa e qualitativa dei dati, con particolare attenzione all'effetto delle caratteristiche distribuzionali sulle dinamiche di memorizzazione e percezione della struttura morfologica. Saranno contrastati i risultati dell'addestramento controllato, senza le frequenze da corpus, per evidenziare gli effetti strutturali della famiglia/serie e l'addestramento con le frequenze da distribuzione naturale, per valutare invece quanto la *token frequency* possa accentuare o bilanciare certi effetti. Le proprietà distribuzionali, oltre a quelle più strettamente superficiali potrebbero influenzare la relazione tra la base e la parola suffissata, portando le basi a richiedere una maggiore attivazione dei suffissi per emergere.

Il Capitolo 5 è focalizzato su una discussione delle evidenze ottenute. Da queste sarà possibile trarre alcune conclusioni che possano confermare le intuizioni circa l'architettura neuro-cognitiva del linguaggio, specificamente riguardante le strutture morfologiche e la loro complessità. È lampante, comunque, che un modello sia sempre una semplificazione, in quanto per definizione è in grado di riprodurre solo alcuni tratti di un fenomeno, escludendone altri, ma esiste una crescente consapevolezza che possa contribuire notevolmente a una precisa caratterizzazione e valutazione dei modelli funzionali del lessico mentale, anche in base alla condizione di input (Pirrelli 2007). Assunto che molti sistemi reali possono essere modellati come reti neurali artificiali, l'analisi quantitativa dei loro complessi stati di organizzazione consente di approfondire le dinamiche organizzative e funzionali del fenomeno indagato. Nel caso specifico della competenza lessicale dei parlanti, attraverso modelli espliciti di ispirazione biologica, le simulazioni possono aiutare a comprendere la natura delle rappresentazioni lessicali complesse e delle relazioni tra esse. Discutendo le virtù di questo approccio (ma anche i suoi eventuali limiti), si potrà riuscire ad evidenziare quei meccanismi di base che governano la memorizzazione e l'elaborazione delle parole partendo dall'evidenza comportamentale psico-cognitiva e dai vincoli neuro-funzionali nella direzione di un modello esplicito auspicabilmente unitario.

Capitolo 2

INQUADRAMENTO TEORICO

Nell'ambito degli studi sul linguaggio accade di imbattersi in una netta contrapposizione metodologica tra la linguistica teorica e la psicolinguistica, dovuta essenzialmente alla natura sperimentale della seconda rispetto a quella formale e non empirica della prima. Questa contrapposizione è da considerarsi restrittiva nonché erronea, poiché la linguistica – in tutte le sue specifiche - è effettivamente una scienza empirica, che indaga un fenomeno e si basa tanto su dati naturalistici (raccolte di produzioni spontanee tipicamente raccolti in corpora) o sperimentali (tipicamente elicitati in contesti sperimentali controllati e progettati).

Se negli anni '80 la gerarchia di Marr¹ sembrava suggerire che ciascun livello possa essere studiato indipendentemente dagli altri, comprendere un fenomeno come il linguaggio umano nella sua complessità richiede necessariamente l'integrazione delle conoscenze sviluppate a ciascun livello. Viceversa, l'approccio contemporaneo di stampo psicolinguistico computazionale si basa sulle interazioni tra i diversi livelli: ne è un chiaro esempio l'approccio volto a rappresentare unità linguistiche semplici come i morfemi come proprietà emergenti da un processo di auto-organizzazione di parole piene nel lessico mentale (James P. Blevins 2016; Marzi, Ferro e Pirrelli 2019; Pirrelli, Ferro e Marzi 2015).

¹Marr (1982) distingue tre livelli di analisi in base ai quali un fenomeno complesso può essere descritto. Il primo è il livello della teoria computazionale, in cui viene spiegato che cos'è una specifica operazione. Il secondo riguarda il dettaglio della computazione e viene definito livello dell'algoritmo. Il terzo è il livello dell'implementazione, che descrive un algoritmo è per mezzo delle operazioni materiali che un calcolatore esegue.

In questa prospettiva, i modelli computazionali come le mappe neurali ricorrenti sono in grado di simulare proprio quegli effetti di emergenza di strutture complesse a partire da semplici processi di memorizzazione di unità non (preventivamente) strutturate. Contestualmente, la simulazione computazionale dei meccanismi che si ispirano ad alcuni dei funzionamenti di base del cervello umano risponde all'esigenza di un maggiore realismo biologico nell'implementazione dei processi di base dell'elaborazione e memorizzazione, ma è spinta soprattutto dalla necessità di gettare un ponte tra il livello algoritmico e quello implementativo (il secondo e terzo dei livelli teorizzati da Marr).

2.1 Una nuova scienza del linguaggio

«Il compito della psicolinguistica (o psicologia del linguaggio) è scoprire come la conoscenza della lingua è rappresentata nella mente (nel cervello) di un parlante, come quest'informazione è normalmente utilizzata nella produzione e comprensione della lingua, e come i parlanti acquisiscono queste abilità» (Akmajian 1979).

Il tentativo di comprendere e spiegare i principi strutturali e i meccanismi cognitivi che sono alla base dell'elaborazione del linguaggio, si è rivelato uno notevole sforzo scientifico che nel tempo ha portato all'integrazione interdisciplinare di diversi domini e metodologie d'indagine. La questione della relazione tra mente e cervello, infatti, ha affascinato scienziati e filosofi fin dalle origini della cultura occidentale, imponendosi come studio centrale del soggetto umano; ciò ha permesso lo sviluppo di diverse teorie, alcune delle quali - concordanti sull'idea che l'attività cerebrale sia intimamente connessa con i processi e le funzioni mentali (Vallar 2020) - sono riproposte nel dibattito attuale. Naturalmente, con il progredire delle conoscenze, della comunità scientifica le riflessioni, le osservazioni e di conseguenza anche le domande, sono andate gradualmente ampliandosi fino a configurarsi, a metà del XIX secolo, in un quadro teorico di riferimento che prevedesse un'indagine dedicata specificamente al linguaggio nell'ambito dei processi mentali. Prima degli anni '50 infatti per gli psicologi il linguaggio non costituiva un fenomeno a sé stante, ma si considerava piuttosto «nient'altro che» il risultato dell'apprendimento, dell'associazione, dell'imitazione. Era solo un comportamento verbale (Bruner e Watson 1983). Dagli anni '50 in poi, il prevalere di un'impostazione

fortemente positivista ed associazionistica, oltretutto ben conforme alla psicologia comportamentista (Skinner 1957) e la linguistica strutturale dell'epoca, fu messo in profonda crisi dalla nascita e l'affermarsi di una nuova scienza del linguaggio, la psicolinguistica, di cui si è soliti collocare l'investitura ufficiale nel 1951, durante un convegno di studi tenutosi all'Università dell'Indiana, negli Stati Uniti, a cui parteciparono psicologi, linguisti, antropologi e informatici². L'ambizione di questa scienza giovane era quella di descrivere i processi psicologici che hanno luogo quando ci serviamo di enunciati (Miller 1967 ponendo le basi sperimentali per le successive ipotesi, per cui alle regole costitutive del sistema linguistico potessero corrispondere altrettanti meccanismi psicologici. Questa presa di coscienza portò ad una prima grande rottura degli schemi comportamentisti che si ebbe a partire dal 1957, quando Noam Chomsky con *Syntactic Structures* (1957) e *Review of verbal behavior* (1959), con estrema lucidità epistemologica diede l'avvio ad una delle grandi rivoluzioni della linguistica, scrivendo pagine definitive contro l'ipotesi che il linguaggio fosse un comportamento imitativo. I suoi assunti fondamentali dichiaravano che: il linguaggio è una competenza, quindi un insieme di regole e meccanismi nella mente di chi parla e comprende; vi sono principi linguistici universali, radicati nel patrimonio genetico della nostra specie; il compito della linguistica è quello di costruire modelli formali che descrivano quei meccanismi.

Anche se oggi la posizione di "radicamento genetico" è quasi del tutto rifiutata, va riconosciuto che quello di Chomsky fu soltanto il trampolino di lancio verso una serie di intuizioni che hanno costellato gli ultimi decenni, dando luogo a un corpus crescente di ricerche basate su nuovi modelli psicolinguistici e tecniche di indagine sempre più moderne.

Se, infatti, l'interesse contemporaneo per l'acquisizione delle conoscenze morfologiche e il modo in cui le parole complesse vengono memorizzate, percepite e prodotte rivela la grammatica generativo-trasformativa come un modello poco funzionale all'analisi dei processi di produzione e percezione, è innegabile l'importanza storica delle evoluzioni e contrapposizioni che all'argomento Chomskiano sono state fatte, anche per voce di autori tutt'altro che ostili a Chomsky (Pullum e Scholz 2002; Jackendoff 2002), circa la scarsa solidità delle evidenze empiriche le quali portano ad una marginalizzazione di importanti variabili sulle questioni della comprensione e della memoria, che attualmente stanno centralizzando il dibattito neu-

²Cfr. Atti del convegno, Osgood e Sebeok 1954.

rolinguistico e psicolinguistico. A questo proposito, una cospicua produzione di evidenze sui meccanismi che regolano questi processi a livello lessicale ha messo in discussione i modelli tradizionali dell'architettura del linguaggio basati sull'ipotesi di una corrispondenza diretta tra componenti modulari della competenza grammaticale, correlati di elaborazione (memoria e computazione) e localizzazioni neuro anatomiche.

2.2 I modelli simbolici della teoria rappresentazionale

Dalla seconda metà del Novecento si è sviluppata sostanzialmente l'idea che la mente umana operi sulla base di rappresentazioni che mediano fra stati di sistemi intelligenti contenenti informazioni³. Le prospettive teoriche che inquadrano il cervello come un sistema che elabora informazioni e che opera sulla base di computazioni (Gazzaniga et al. 2015) non sempre risultano compatibili, orientandosi in prominenza o dell'aspetto cognitivo o dell'aspetto linguistico, quindi privilegiando ora una visione funzionalista, ora una visione strettamente biologica delle facoltà mentali⁴. Proprio nel pieno degli anni Ottanta dello scorso secolo il dibattito si accendeva fra queste due preponderanti correnti di pensiero che interpretano in maniera differente il rapporto fra linguaggio e funzioni cerebrali:

- il modularismo, di stampo cognitivista, che sostiene la modularità della mente e l'autonomia del sistema linguistico.
- il connessionismo, di stampo empirista e anti-innatista, che fa riferimento a una generica facoltà di categorizzazione comune all'uomo e ad altri esseri viventi.

Una caratteristica fondamentale dell'approccio MODULARISTA è l'organizzazione gerarchico-sequenziale delle strutture di elaborazione, che deriva dalla assunzione secondo cui la struttura e il funzionamento del calcolatore di Von Neumann sarebbe un modello plausibile della architettura e del funzionamento della mente (Ulam 1958).

³Queste rappresentazioni dei sistemi cognitivi sono caratterizzate dal fatto che sono permanenti nel tempo e che possano essere tipiche dei vari sistemi percettivi, o eventualmente amodali cfr. Markman e Dietrich 2000.

⁴Per i teorici del funzionalismo (cfr. Putnam 1987; Putnam 1993; Fodor J. 1988; J. A. Fodor 1990.) l'obiezione ad assunti strettamente materialistici è mossa verso l'impossibilità di spiegare una visione soggettiva della realtà esterna.

Così come per un computer ogni calcolo non è altro che il risultato della esecuzione di un algoritmo, anche per il funzionamento della mente si è assunto che l'elaborazione delle informazioni sia rappresentabile come un flusso sequenziale di informazioni in input che a partire dagli organi di senso giungono ai meccanismi decisionali centrali da cui poi restituiscono un output, senza postulare alcun isomorfismo tra i processi cognitivi e il cervello. La tassonomia funzionale dei processi psicologici introdotta da Fodor (1988) distingue tre livelli, per cui il flusso delle informazioni in entrata accede a tali meccanismi nell'ordine così enucleato: trasduttori, sistemi di input e processi centrali. I sistemi di input, che trattano informazioni giunte dall'apparato percettivo, sono modulari, pertanto condividono tutte le proprietà tipiche delle facoltà verticali⁵ e sono caratterizzati da specificità per dominio, senza possibilità di interagire. La modularità altresì attribuita agli analizzatori di input presuppone che le rappresentazioni risultanti debbano interfacciarsi, di conseguenza devono esistere dei sistemi cognitivi non modulari che trattano informazioni da domini differenziati. I sistemi centrali integrano le conoscenze specifiche e sono responsabili delle funzioni cognitive superiori (figura2.1).

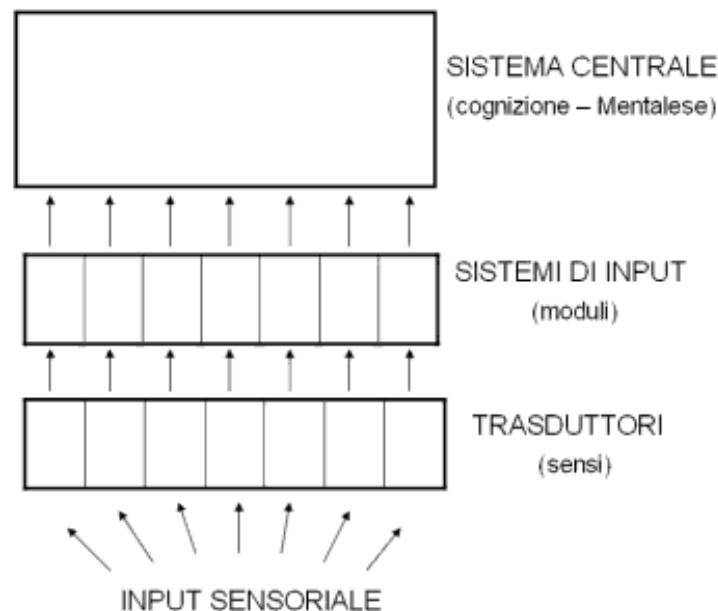


Figura 2.1: Rappresentazione schematica dell'architettura cognitiva della mente modulare ipotizzata da Jerry A. Fodor (tratto da Galloni, 2006).

⁵La struttura verticale è da intendersi in maniera sinonimica di quella modulare.

In questa idea architettonica della mente, l'analisi percettiva non è propriamente associata al pensiero (Fodor J. 1988), ma è piuttosto lo specchio di un sistema computazionale che media tra output dei trasduttori e meccanismi cognitivi centrali per «ottenere informazioni sul mondo in un formato idoneo per l'accesso a quei processi centrali che mediano la fissazione delle credenze» (Fodor J. 1988 :80). L'ipotesi innatista, secondo la quale gran parte della struttura del linguaggio è trasmessa geneticamente grazie al carattere ereditario della struttura cerebrale, lascia il posto alle teorie modulari secondo cui quanto più il cervello è costituito da parti specializzate, tanto più è probabile che la loro specializzazione si sviluppi mediante forme di apprendimento tipicamente tradotte in meccanismi di stimolo-risposta.

In maniera espressamente antitetica alla teoria modulare, i teorici del CONNESSIONISMO ritengono quest'ultima una teoria alquanto inadeguata rispetto al fatto evidente che il sistema nervoso è costituito da una rete di elementi fortemente interconnessi tra di loro. Questo specifico orientamento di ricerca è rivolto per lo più allo studio dei processi cognitivi attraverso programmi di elaborazione che, essendo appunto dotati di una struttura analoga a quella delle reti neurali, sono in grado di simulare a livello computazionale il funzionamento del sistema nervoso. Sotto l'egida del connessionismo la simulazione non è più soltanto descrittiva, ma diventa esplicativa quanto più approssima il comportamento linguistico consentendo, inoltre, di controllare perfettamente le variabili e compiere manipolazioni che sarebbero impossibili nella realtà. Con la pubblicazione del manifesto *Parallel Distributed Processing* James L McClelland, David E Rumelhart, Group et al. 1986, è stata capovolta in modo radicale la maggior parte delle assunzioni del cognitivismo, il quale nei decenni precedenti aveva definito i criteri di studio delle architetture mentali in base al binomio regola-rappresentazione. Il connessionismo per contro è un approccio prettamente simulativo ispirato dalla analogia che intercorre tra reti neurali artificiali e il funzionamento del cervello umano, nell'ipotesi che i processi cognitivi possono essere descritti e simulati efficacemente solo da una macchina la cui struttura e il cui funzionamento siano simili a quelli del cervello. Secondo Rumelhart e McClelland (1986) – ma ciò è riconosciuto anche da chi si contrappone all'approccio connessionista (Steven Pinker 1994) - i processi cognitivi non possono ridursi a moduli separati della mente, né funzionare in modo seriale, ma piuttosto in modo parallelo. Non vi è, dunque, un substrato simbolico, ma solo connessioni tra nodi all'interno di una rete neurale, ovvero un sistema di elaborazione di

informazioni, composto da un elevato numero di unità – i neuroni (artificiali). Ogni neurone compie sempre e soltanto una semplice operazione matematica, mentre l'unico elemento variabile della rete è l'intensità delle connessioni tra i diversi nodi, che viene calcolata durante la fase di apprendimento. L'informazione non è elaborata in modo sequenziale attraverso l'applicazione ordinata di regole, ma consiste nell'attivazione simultanea, e quindi «parallela», di un numero elevato di nodi della rete. Questa intuizione si evidenzia soprattutto nei brevissimi tempi che intercorrono fra un input esterno e la risposta della rete ad esso. Se l'informazione venisse gestita in modo seriale dai neuroni richiederebbe troppo tempo, per cui è naturale assumere che i nostri processi cognitivi debbano prevedere una distribuzione ed una gestione parallela delle informazioni a livello cerebrale. A partire da questi modelli simulativi di apprendimento lessicale si ha un allontanamento sempre più visibile dal paradigma simbolico, anche su un certo piano di astrazione, introducendo il concetto di conoscenza distribuita in cui gli elementi che la compongono sono elementi sub-simbolici, caratterizzati da proprietà matematiche e senza uno specifico valore semantico. Anche il concetto di rappresentazione viene reinterpretato alla luce della struttura del modello all'interno della quale è costruita: essa non è più identificata in un'unità semantica, un simbolo, ma è intesa come un processo di attivazione di molte unità contemporaneamente senza che nessuna di esse abbia un proprio significato. A questo proposito, secondo Rumelhart il cervello umano è composto da elementi che, in numero elevatissimo (miliardi di neuroni), elaborano le informazioni in modo cooperativo e parallelo, per cui l'apprendimento stesso consiste essenzialmente nel rafforzamento o nell'inibizione di alcune connessioni a seconda dell'input. Una rete neurale connessionista ha almeno tre strati di unità: un livello di elaborazione delle unità di input, le quali, attraverso connessioni ad un livello nascosto, sono connesse alle unità di output (Figura 2.2).

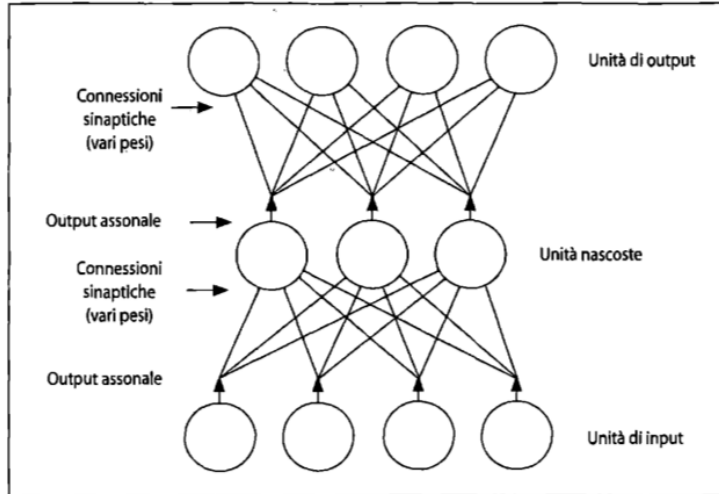


Figura 2.2:

Esempio di rete connessionista.

(tratto da Cacciari 2011, adattamento da Churchland 1989).

Quando un output non coincide con l'obiettivo desiderato/atteso, è calcolato l'errore utilizzato a sua volta per modificare a ritroso i pesi delle connessioni, in modo da indurre l'output atteso. Dopo un certo numero di cicli di addestramento – e di correzione supervisionata dei pesi delle connessioni - la rete sarà in grado di rispondere correttamente all'input. Ciò che la rete ha appreso in questo modo ha una rappresentazione sub-simbolica, come assetto globale dei pesi. Tale funzionamento di regolazione dei pesi tra nodi/neuroni vuole essere in linea con la regola di Hebb⁶, secondo la quale se due neuroni collegati tra loro sono contestualmente attivi, l'efficacia sinaptica della connessione viene rinforzata e in seguito a questo rafforzamento, la sola attivazione del nodo presinaptico è sufficiente a causare l'attivazione del nodo postsinaptico. Tale associazione verrà ulteriormente rinforzata ogni volta che i due nodi si attiveranno sequenzialmente. In un'analogia computazionale, all'inizio i pesi delle connessioni della rete sono tutti a zero, poi per ciascun pattern di attivazione, i nodi della rete vengono attivati ai valori bipolari di quel pattern ed i pesi vengono modificati tramite la regola hebbiana.

Secondo questo modello formale l'apprendimento è spiegabile secondo tre ipotesi (Florea-no e Mattiussi 2002), ciascuna delle quali ha col tempo ricevuto adeguate conferme neuro-

⁶Nel 1949 lo psicologo canadese Donald Hebb affermò che le leggi del condizionamento classico riflettevano le proprietà di funzionamento degli elementi costitutivi del sistema nervoso.

funzionali grazie all'avvento delle tecniche di *neuroimaging*. La prima ipotesi è che i neuroni corticali rafforzino le loro connessioni quando risultano con frequenza attivi contemporaneamente, ovvero quando si verifica una correlazione della loro attivazione. Questo principio di apprendimento associativo sembra essere valido per la maggioranza dei neuroni corticali; a esso si fa comunemente riferimento, appunto, come regola di Hebb (1949). La seconda ipotesi è che la corteccia sia un'enorme memoria associativa in cui il rafforzamento delle sinapsi abbia luogo non solo tra neuroni vicini, ma anche tra neuroni in aree corticali distanti. Secondo la terza ipotesi, la contemporanea e frequente attivazione di un gruppo di neuroni che dà luogo al rafforzamento sinaptico ha conseguenze funzionali importanti: i neuroni fortemente connessi probabilmente agiscono insieme, come un'unità funzionale. Se vengono attivati solo alcuni dei neuroni, si attiverà l'intero gruppo, a causa delle forti connessioni tra i membri del gruppo stesso. Gli insiemi hebbiani di cellule si possono definire come unità funzionali composte da molti neuroni che si formano in una rete associativa, la corteccia, come risultato di una frequente attività neuronale simultanea che causa un rafforzamento sinaptico. L'abilità di ricordare uno stimolo, una sequenza di stimoli o un evento, dipende proprio dai processi che modificano continuamente i nostri circuiti neurali. Il meccanismo di base è la plasticità sinaptica⁷ che, secondo la metafora dell'interruttore sinaptico, è descritta come un cambiamento permanente, inteso come potenziamento a lungo termine, che interviene all'interno di un insieme di sinapsi in conseguenza di un'attività neurale evocata da un'esperienza sensoriale. Lo stesso meccanismo può portare a un rapido decadimento delle tracce di memoria, a causa del flusso continuo di nuove esperienze che tendono a generare nuove memorie. Le idee di Hebb sono state espresse più sotto forma di speculazioni sul funzionamento neurale che sotto forma di modelli di elaborazione concreti, ma il suo pensiero cattura parte delle intuizioni dei meccanismi di elaborazione paralleli distribuiti, costituendo le fondamenta su cui si basano o derivano tutti gli algoritmi di apprendimento dedicati agli studi sul linguaggio.

La critica maggiore che può essere mossa all'approccio connessionista di matrice Rume-

⁷Questa proprietà permette al sistema nervoso di modificare la sua struttura e la sua funzionalità in modo più o meno duraturo e dipendente dagli eventi che li influenzano come ad esempio l'esperienza. In particolare è la capacità del sistema nervoso di modificare l'intensità delle relazioni inter-neuronali (sinapsi), di instaurarne di nuove (sinaptogenesi) e di eliminarne alcune (*pruning*).

lhartiana risiede proprio nella correzione SUPERVISIONATA dei “legami hebbiani tra i nodi della rete. Lo stesso Hebb – qualche decennio prima – sottolineava l’importanza dell’input e della reiterazione di esso per il rafforzamento delle connessioni tra neuroni: rafforzamento che è il frutto di una risposta adattiva più che di una supervisione al fine del raggiungimento di un output desiderato. L’approccio connessionista è stato vivacemente criticato da Pinker e Prince (1988), i quali hanno evidenziato alcuni limiti del modello, quali, ad esempio, riconducibili all’interpretazione evolutiva degli errori, e all’incapacità del modello di produrre alcune forme di iper-regolarizzazione dei verbi irregolari, o il caso degli allotropi (possibili forme del *past tense* sia di stampo regolare che irregolare, come nel caso dell’italiano *veduto* e *visto*). Pinker e Prince ritengono che il ricorso a sistemi di regole per la spiegazione del linguaggio umano e del suo sviluppo è funzionale e non sostituibile da modelli connessionisti.

2.2.1 L’approccio duale

Poiché molti studi psicolinguistici hanno affrontato il problema del riconoscimento di parole morfologicamente complesse, è stata riservata particolare attenzione alle condizioni che ne determinerebbero l’elaborazione per decomposizione in morfemi, o per recupero dell’intera forma. In particolare, il dibattito verte su quale tipo di informazione sia soggetta a memorizzazione e quale sia invece soggetta a regole lessicali nel combinare le unità di conoscenza memorizzata. L’approccio duale si configura in una famiglia di modelli psicolinguistici che include due vie parallele di accesso al lessico: una via più veloce per cui le forme di parole morfologicamente complesse possono essere elaborate in modo associativo, cioè attraverso rappresentazioni memorizzate in forma completa, e una via più lenta, basata su regole che scompongono o analizzano le forme di parole flesse in costituenti morfologici⁸.

⁸Nonostante tali modelli sostengano meccanismi di rappresentazione nettamente distinti per verbi regolari e irregolari, alcuni studi in letteratura contemplano l’esistenza di famiglie di verbi “sub-regolari”, che condividono caratteristiche morfo-fonologiche e seguono lo stesso pattern flessivo. Il dibattito sui pattern sub-regolari si è sviluppato soprattutto attorno a lingue come l’italiano, basate sulla suddivisione in classi flessive, ognuna caratterizzata da uno specifico paradigma regolare, e su una mutevole aggregazione di famiglie sub-regolari (Cfr. Clahsen 2006).

Pinker e Ullman (2002) in particolare difendono la teoria secondo cui le forme irregolari sono memorizzate nel lessico, mentre le forme regolari possono essere prodotte da una regola di concatenazione, che richiede il sistema procedurale. Peculiarità dei modelli duali (S. Pinker e Prince 1988) è che sono essenzialmente basati sulla distinzione tra il lessico mentale e una grammatica mentale, ovvero un insieme di regole ricorsive e combinatorie che consentono al parlante di creare parole flesse, derivate o nuove. Questa opposizione implica una separazione tra il sistema di memoria associativa coinvolto nella memorizzazione e nel recupero di forme irregolari e le procedure combinatorie impiegate per elaborare le forme regolari. Secondo questi modelli, le parole irregolari sono riconosciute come unità linguistiche non scomponibili, e quindi registrate come forme intere. Al contrario, le parole regolari vengono elaborate attraverso il loro morfemi, ognuno dei quali viene recuperato dalla memoria.

A questo approccio dicotomico sono stati proposti modelli "misti" tra cui l'AAM (*Augmented Addressed Morphology*, A. Caramazza, A. Laudanna e Romani 1988), secondo il quale una scomposizione morfologica è più probabile per parole o neologismi poco frequenti che non sono rappresentati nel lessico mentale, e il Race Model (Schreuder e R. H. Baayen 1995), per cui entrambi i percorsi sono attivati in parallelo per tutte le parole, e che il successo di un percorso rispetto all'altro è legato prevalentemente alla frequenza della parola elaborata più che alla sua regolarità morfologica.

2.3 La rappresentazione dell'informazione

I modelli computazionali più recenti, nei quali si riconosce anche la comunità psicolinguistica, sono quelli focalizzati sui processi morfologici, a livello di riconoscimento delle parole, cui sottendono i concetti fondamentali di lessico mentale (Oldfield 1966) e di entrata lessicale. Premesso che l'acquisizione lessicale è resa possibile dal concomitante sviluppo delle funzioni mnestiche, si può definire il lessico mentale, nel suo senso più generico e comune, come il sistema della memoria a lungo termine di immagazzinare, attraverso il tempo, la conoscenza di un numero più o meno vasto di parole⁹.

⁹Di tali parole i parlanti immagazzinano informazioni di vario tipo, relative all'ortografia, alla struttura fonologica, alla struttura morfologica, alle caratteristiche sintattiche, oltre a una serie di informazioni semantiche,

Le informazioni stesse riguardanti le parole sono immagazzinate sotto forma di entrate lessicali, con possibili collegamenti tra loro tramite la condivisione di certe caratteristiche (di natura fonologica, morfologica, morfosintattica o semantica)¹⁰.

Nella modellazione di parole complesse nella memoria lessicale del parlante è presupposta la capacità di conservare sequenze di unità simboliche nella memoria a lungo termine, in modo che siano sempre accessibili e recuperabili per l'elaborazione, tramite meccanismi di accesso al lessico, inteso come «quel processo in cui tutte le informazioni che concernono una parola – di tipo fonologico, morfologico, semantico e sintattico – sono contemporaneamente disponibili, senza che ciò necessariamente implichi che sia avvenuta la selezione di una parola specifica» (Cacciari 2011)¹¹. Sebbene molte teorie abbiano adottato la metafora del lessico mentale come un dizionario o magazzino in cui sono conservate tutte le informazioni sulle parole conosciute, altri modelli che rifiutano l'esistenza stessa di questo “magazzino mentale”. In particolare, i modelli distribuiti tipici del connessionismo postulano che alle parole corrispondano pattern di attivazione di diverse unità all'interno di una rete neurale, facendo dello stimolo l'elemento di principale innesco di questo processo. Sul piano cognitivo, comprendere nel dettaglio come avviene il riconoscimento di una parola, significa riuscire descrivere una serie di meccanismi complessi che elaborano l'informazione. La rappresentazione percettiva che segue quest'elaborazione, stabilisce un contatto con una o più forme all'interno del lessico mentale, in un insieme organizzato di unità di rappresentazione corrispondenti a parole che si attivano in risposta ad uno stimolo sensoriale. Questi processi percettivi sono stati al centro di vivaci discussioni da parte dei connessionisti; la questione, di fatto, è come venga riconosciuta una parola, o ancora più specificamente di come vengono rappresentate e quindi elaborate le parole morfologicamente complesse nel lessico.

quali il significato letterale, idiomatico, pragmatico, culturale, ma non necessariamente in modo completo.

¹⁰La maggior parte delle teorie sul linguaggio concordano sul fatto che i parlanti possiedono un lessico mentale, che usano sia nella comprensione sia nella produzione del linguaggio: riconosciamo una parola che abbiamo ascoltato o letto, in quanto essa corrisponde a un'entrata del lessico mentale (cioè ad una voce, ossia un particolare vocabolo, presente in questo lessico); allo stesso modo, nel produrre un frase, a voce o scritta, estraiano dal nostro lessico mentale la parola che corrisponde a quel che vogliamo comunicare. La presenza di queste relazioni è testimoniata dal fatto che la velocità e l'accuratezza di accesso alle parole nel lessico mentale è influenzata dalla loro vicinanza.

¹¹Un principio soggiacente a questo tipo di processo segue i livelli di attivazione e decadimento dell'attivazione.

Semplificando, se l'esistenza di un lessico mentale trova d'accordo la maggior parte degli psicolinguisti, l'oggetto del contendere è se le rappresentazioni relative alle diverse entrate lessicali siano basate su parole intere o morfologicamente decomposte e quindi su come avviene – di fatto - l'accesso alle diverse informazioni contenute nel lessico (Balota, d'Arcais e Rayner 1990; Taylor e J. Fodor 1990; Marslen-Wilson et al. 1994). Il maggior dibattito è sulla natura dell'unità lessicale di riferimento, e sull'accesso al lessico che può essere *word-based* o *morpheme-based* (cfr § 2.2.1). In ogni caso, a prescindere dal modello di riferimento, postulare l'esistenza di una serie di criteri di organizzazione dell'informazione lessicale è di per sé un passo fondamentale, poiché il lessico stesso necessita di una struttura per funzionare in maniera economica ed efficace. In particolare, è fondamentale comprendere come la conoscenza morfologica è rappresentata nella mente umana e come viene utilizzata nell'elaborazione del linguaggio, presupponendo che la conoscenza in un particolare dominio della cognizione umana comprende sempre sia l'immagazzinamento di informazioni sia la capacità di produrne di nuove.

2.3.1 L'informazione morfologica

In ambito psicolinguistico è opportuno pensare alla morfologia¹² non come al semplice studio delle parole, ma come allo studio delle relazioni, della formazione e della percezione delle parole, ovvero ponendo l'accento sia sui legami semantico-formali che ogni parola può avere con altre parole della stessa lingua, sia su relazioni di natura puramente superficiale. Si tratta essenzialmente dello studio delle implicazioni contestuali che le parole stesse determinano da un punto di vista sintattico e semantico, così come di quei principi dinamici che ne controllano la formazione interna. Molte ricerche, di conseguenza, sono state dedicate a verificare se nel lessico mentale esista effettivamente un livello morfologico di rappresentazione usato nella comprensione delle parole per dividerle nei morfemi componenti, e se la struttura del lessico mentale sia basata su parole intere oppure su unità morfemiche (nel dibattito che contrappone un modello del lessico mentale *word-based* e uno *morpheme-based*).

¹²Sulla morfologia teorica si vedano, ad esempio, Spencer 1991; Spencer e Zwicky 2001; Booij et al. 2000.

- L'approccio morpheme-based prevede che le parole morfologicamente complesse siano rappresentate in forma decomposta e che le parole morfologicamente relate abbiano unità morfologiche condivise. Ad esempio, le parole morfologicamente collegate, come: camminare, camminando, camminerò, camminamento, camminatore, ecc. condividono la stessa rappresentazione di base; alla radice cammin, si aggiungono diversi suffissi per formare le forme flesse e derivate: cammin -ando; cammin -erò; cammin -amento, ecc.
- L'approccio word-based vede le parole rappresentate come forme intere, non decomposte, sebbene interconnesse attraverso relazioni di natura morfologica. In questo modello non esiste una rappresentazione indipendente della radice/base nel lessico mentale, ma contempla invece tante voci quante sono le forme flesse o derivate di una parola: camminare, camminando, camminerò, camminamento, camminatore, ecc.

Il ruolo del morfema è stato estremamente influente nella psicologia e nelle scienze cognitive, dove è opinione diffusa che i morfemi debbano esistere in qualche maniera in una rappresentazione della mente. In particolare, le problematiche legate alla modalità di accesso al lessico sono state a lungo indagate.

In un'impostazione specificamente modularista, il processo di accesso al lessico avverrebbe in due distinte fasi: il livello del recupero del lemma, che produce solamente l'attivazione della specifica entrata lessicale, e il livello del recupero del lessema in cui avvengono le integrazioni della parola con le informazioni del contesto in cui è inserita¹³. Secondo tale approccio, ne deriva che la struttura morfologica di una parola è combinatoria e che i suoi elementi costituenti hanno entrate indipendenti nel lessico mentale; una regola deterministica combinerebbe questi morfemi nella loro struttura complessa e deriverebbe composizionalmente il suo significato dai significati delle parti, ma senza che le forme complesse vengano immagazzinate in memoria. Il modello dual-route, o Dichiarativo-Procedurale (S. Pinker e Prince 1988; Pinker e Ullman 2002, §2.1.2), postula due sistemi di elaborazione cognitiva fundamentalmente distinti:

¹³Il lessico mentale è considerato principalmente un deposito di conoscenze riguardo alle parole di una data lingua; al suo interno ogni entrata lessicale consiste di due parti: il lemma, che contiene informazioni semantiche inerenti la parola, e informazioni che riguardano le caratteristiche grammaticali riguardanti la parola stessa, quali per esempio, la categoria grammaticale (nome, verbo ecc.), e il genere grammaticale (maschile, femminile, o neutra, per le lingue che lo possiedono); e il lessema, che contiene la struttura fonologica e/o ortografica.

uno è basato su regole e si applica alle parole flesse regolari, mentre l'altro che è basato su meccanismi associativi, si applica a parole complesse irregolari. Per contro, vari modelli associativi di stampo connessionista sostengono che tutte le parole, semplici o complesse, regolari o irregolari, vengono elaborate fundamentalmente allo stesso modo (Mark S. Seidenberg e James L. McClelland 1989). In tali modelli, gli effetti morfologici non emergono come risultato di rappresentazioni morfemiche, ma in virtù delle relazioni di forma e significato che la struttura morfologica codifica. In generale la morfologia derivativa pone problematiche di una certa rilevanza, con un quadro sicuramente più complesso, soprattutto tenendo conto che le forme del tutto regolari sono più rare per derivazione che per inflessione (Plag e Balling 2020). In base all'approccio duale, tipicamente solo le forme irregolari sono memorizzate nel lessico come voci intere. Questo approccio "costruttivo" della competenza morfologica (in contrapposizione alla competenza "astrattiva" definita da Blevins 2006, per un approccio alla struttura delle parole che prevede un lessico ridondante) dove basi e affissi sono gli elementi costitutivi (costruttivi, appunto) della competenza morfologica in un lessico privo di ridondanza. L'approccio astrattivo vede, invece, nelle forme piene i mattoni della competenza morfologica, e le parti sub-lessicali ricorrenti (i morfemi) come "astrazioni" sulle forme morfologicamente complete. Ciò che maggiormente risulta importante per le finalità del presente lavoro, riguarda il dibattito teorico-sperimentale sui concetti di memorizzazione ed elaborazione. L'esistenza di una netta dicotomia tra questi processi è stata fortemente messa in dubbio dagli approcci all'organizzazione del lessico morfologico (Joan L Bybee 1985, 1995, 2010), che definiscono un lessico mentale ridondante, la cui organizzazione e struttura siano funzionali e ottimizzati all'accesso e al recupero lessicale, dato che i normali processi di produzione e comprensione dell'input linguistico devono operare entro vincoli temporali di poche centinaia di millisecondi. Basta pensare a questo per capire come la conoscenza lessicale non possa essere organizzata per minimizzare la memoria o rispondere a necessità logiche, ma che sia, invece, funzionalmente strutturata per massimizzare l'elaborazione (G. Libben 2005). Nella letteratura psicolinguistica, questa consapevolezza ha portato a un'ampia rivalutazione del ruolo dei morfemi sul ruolo delle unità sublessicali nell'elaborazione lessicale. Recenti evidenze sperimentali sull'elaborazione morfologica hanno suggerito una visione meno deterministica e modulare della natura e dell'interazione tra la conoscenza e l'elaborazione delle parole memorizzate.

2.4 Evidenza comportamentale

Benché sia noto che le parole sono in qualche maniera associate tra loro in termini di ortografia, fonologia, morfologia o semantica, la natura precisa di queste relazioni è stata ampiamente dibattuta. Alla domanda se le rappresentazioni lessicali siano basate su parole piene o siano organizzate sulla base di relazioni morfologiche, viene tipicamente data risposta progettando appropriati esperimenti di decisione lessicale con o senza *priming* (Marslen-Wilson et al. 1994). Un minuzioso campionamento della letteratura sperimentale permette di utilizzare il compito di decisione lessicale — anche con tecniche di priming mascherato e non mascherato in funzione dei tempi di esposizione allo stesso — per indagare la rappresentazione delle parole morfologicamente complesse. Negli studi comportamentali, i compiti di decisione lessicale consistono nel decidere se un certo stimolo appartiene o meno alla lingua di riferimento e possono essere effettuati con stimoli presentati visivamente o in maniera cross-modale¹⁴. Questo genere di esperimento viene progettato sottoponendo al soggetto parole e pseudo-parole, che tuttavia rispettano i vincoli ortografici e fonologici della lingua¹⁵; nello svolgimento del task sono registrati i tempi di risposta (o *reaction times*, RT) e il numero di risposte giuste (accuratezza), per cui a tempi di risposta differenti corrispondono tempi di elaborazione diversi e quindi processi cognitivi diversi. La latenza di risposta, ovvero il tempo necessario per accedere a una parola nel lessico e per accertare che la voce lessicale presente nel lessico mentale corrisponda alla parola che è stata presentata (Cfr. Rueckl e Galantucci 2005), insieme all'accuratezza sembra essere un aspetto fortemente determinato dalle caratteristiche degli stimoli. Tra le più influenti vi sono, senza dubbio, la frequenza delle parole, il numero e la frequenza dei vicini ortografici e altre caratteristiche delle parole come la frequenza delle lettere o dei bigrammi/trigrammi, la concretezza o l'immaginabilità (Mark S Seidenberg 1995). Poiché l'evidenza neuro-funzionale porta a una visione della competenza lessicale come sistema adattivo è natu-

¹⁴Condizioni di presentazione particolarmente ben esplorate per il prime includono manipolazioni della durata dell'esposizione del prime e la presenza o assenza di un pattern mascherato che lo precede. Nella scelta cross-modale il soggetto ascolta una parola, mentre sullo schermo appare uno stimolo su cui effettuare la decisione lessicale.

¹⁵Le decisioni vengono comunicate premendo un pulsante della tastiera o del mouse o tramite uno speciale dispositivo chiamato pulsantiera collegato al computer.

rale assumere che, mentre i principi cognitivi - non specificamente linguistici- interagiscono con memoria a breve termine e memoria a lungo termine, i fattori distribuzionali (tipicamente *token* e *type frequency*) influenzano fortemente la percezione, modulando gli aspetti competitivi in termini di agevolazione o inibizione sul riconoscimento; peraltro la frequenza di una parola confrontata con la frequenza degli altri membri della stessa famiglia lessicale, influenza la capacità del parlante di percepirne la struttura (Moscoso del Prado Martin et al. 2004)¹⁶.

Nei compiti di decisione lessicale, la misurazione dei tempi di reazione in particolare maniera offre l'evidenza sperimentale degli effetti di trasferimento di attivazione dalla parola *prime* alla parola *target*, con un esito facilitatorio sul riconoscimento. Assumendo che la relazioni superficiali tra le parole abbiano un effetto sulla loro elaborazione, grazie agli esperimenti di *lexical decision*, specificamente con *priming*, emerge il ruolo rilevante della struttura sublessicale¹⁷ per cui gli effetti di relazione della struttura morfologica più superficiale risulta mostrare i primi effetti sull'accesso (sub)lessicale, con un effetto di attivazione delle proprietà semantiche solo successivo.

2.5 Modelli computazionali del lessico mentale

Lo scopo di un modello è essenzialmente quello di aiutare a decomporre un processo complesso in un insieme ristretto di sotto-processi interagenti, in condizioni controllate. In linea di principio esistono infiniti modelli in grado di replicare lo stesso compito, ma un aspetto importante per la sua validazione è il confronto tra i risultati prodotti da un modello e quelli di altri modelli, purché sullo stesso compito e possibilmente sullo stesso insieme di dati di test. I modelli computazionali del lessico mentale si basano chiaramente sulle maggiori teorie della rappresentazione lessicale, nel tentativo di rendere espliciti meccanismi e condizioni che le teorie spesso danno per scontati (Ferro, Pezzulo e Pirrelli 2010). Come già accennato, alcuni ap-

¹⁶Una famiglia lessicale con una forma ad alta frequenza viene acquisita più facilmente, rispetto alla generalizzazione verso nuove forme. Al contrario, una famiglia lessicale più ampia, i cui membri presentano una distribuzione uniforme di frequenze portano ad un'acquisizione competitiva rallentata, ma favorendo la generalizzazione.

¹⁷Maggiore la capacità del parlante di individuare strutture ricorrenti, maggiore sarà la capacità di elaborare una certa quantità di dati in input.

procci computazionali prevedono che ogni passo dell'elaborazione sia predefinito nei dettagli e descritto in forma esplicita nella fase di programmazione; le regole rispettano rigorosamente la loro forma, escludendo ogni tipo di flessibilità: la non soddisfazione di una sola delle condizioni implica la sospensione dell'intero processo. Per altri approcci invece, il discorso della plausibilità biologica diventa un argomento essenziale, che non può essere eluso nella pretesa di modellazione. E proprio il tentativo di fornire una modellazione neuro-computazionale del lessico mentale, porta alla luce una problematica essenziale: se immagazzinare una parola non significa semplicemente conservarla nella memoria, ma richiede un livello di codifica e la manipolazione attraverso strutture topologiche che si adattano in modo ottimale alla produzione e alla comprensione delle parole stesse, non è altrettanto intuitiva la posizione da assumere sulle dinamiche di accesso lessicale. Gli approcci interdisciplinari che sfumano i confini tra psicologia, neuroscienze e linguistica hanno portato alla teorizzazione dei diversi approcci a come le parole sono rappresentate, accedute e recuperate dal lessico mentale. Alcune teorie sostengono che la selezione delle entrate lessicali avviene in maniera seriale (modello di rappresentazione/elaborazione di tipo indiretto); altre, invece, sostengono l'idea di un'attivazione in parallelo (modello diretto), fortemente legate al concetto di soglia di attivazione¹⁸. In un modello indiretto come quello di Forster 1976, le parole con relativa informazione sono immagazzinate in un archivio generale (master file) al quale si accederebbe tramite alcuni archivi periferici, dove le parole sono immagazzinate come coppie di codici d'accesso ed indicatori. Gli archivi periferici sono costituiti da un certo numero di bins, che raggruppano le parole per grado di somiglianza.

Un importante modello di attivazione in parallelo (modello diretto) è il *Cohort Model* di Marslen-Wilson (1987, e 1999 nella versione *Distributed Cohort Model*), che si ispira alla metafora dell'unità militare antica dell'esercito romano, chiamata appunto coorte. Questo modello postula l'esistenza di un sistema di riconoscimento delle parole in cui come reazione all'input tipicamente uditivo si attivano tutte le parole del lessico mentale che iniziano con lo stesso

¹⁸La distinzione generale tra modelli indiretti e diretti di rappresentazione ed elaborazione lessicale può essere rappresentata con due metafore facilmente concepibili. Garman 1992 ritrae i modelli diretti come software di elaborazione testi: per trovare una parola nel database del computer è necessario digitarne tante lettere quante sono sufficienti per distinguerla dagli altri elementi memorizzati. D'altra parte, un modello indiretto può essere paragonato alla ricerca di una parola in un dizionario o di un libro in una biblioteca.

gruppo di suoni; queste parole attivate – in parallelo - costituiscono una coorte di candidati al riconoscimento, e man mano che la nuova informazione acustica viene presentata, vengono scartate le parole che non corrispondono più, fino a che rimane la sola parola target. Questo processo presuppone l'esistenza di un "punto di unicità" (*uniqueness point*), identificabile con il punto in cui la parola è riconosciuta come diversa dagli altri membri possibili della coorte e identica al segnale di ingresso¹⁹ (il concetto di *uniqueness point* è stato poi ripreso da Wurm et al. 2006; Marzi, Ferro e Pirrelli 2019, distinguendolo in primo punto di unicità, o UP1, e punto di unicità complesso, o CUP). In questo modello sia frequenza della parola obiettivo sia quella delle parole della coorte rivestono in ruolo determinante per il riconoscimento della parola target. Analogamente, è rilevante il contesto: le informazioni contestuali semantico-pragmatiche sembrano essere rilevanti nel determinare quali membri della coorte vengono disattivati (i.e., non attivati in parallelo) in base alla mancata corrispondenza contestuale, facilitando così l'elaborazione semantica²⁰. Un altro modello diretto da menzionare è il modello logogen di Morton (1969, 1979, 1980), che prevede sì un'attivazione seriale, ma di più unità che operano in parallelo. Questo modello concepisce i cosiddetti demoni, chiamati "logogens", come recettori di parole: ogni parola nota ha una rappresentazione, ovvero un logogen con un suo livello di attivazione (dipendente essenzialmente dalla frequenza) che accumulando proprietà sensoriali di stimoli linguistici, emette poi la propria risposta quando i livelli di attivazione superano una certa soglia valore, necessaria al riconoscimento. Ogni parola scritta attiva più logogen in funzione della loro somiglianza ortografica con la parola target, ma solo uno raggiunge la soglia per l'identificazione. È proprio sulla scia di Morton che la corrente connessionista, con il *Parallel Distributed Processing*, ha rimarcato l'idea che l'elaborazione lessicale sia costituita da una rete di nodi di elaborazione paralleli, portando però alla ribalta un fattore essenziale, mancante in tutti i modelli precedenti: la dinamica temporale dell'apprendimento. In questa prospettiva, il modello connessionista di maggiore impatto, che trae ispirazione dal modello di

¹⁹Singleton (1999) illustra questo processo con la parola elefante [ˈɛlɪfənt]. Spiega che la coorte iniziale della parola [ˈɛli], può includere parole come "element", ma non "elephantine" a causa dell'accento iniziale della parola. Il punto di unicità è presumibilmente il riconoscimento del suono [f] poiché non c'è altra parola nella lingua inglese che inizi con il suono [ˈɛlif].

²⁰Emmorey e Fromkin (1988) citano prove sia a favore che contro il fatto che l'inizio delle parole è responsabile della facilità di accesso, ma aggiungono che le loro implicazioni per il modello di coorte non sono ancora chiare.

coorte di Marslen-Wilson, è il TRACE Model (James L McClelland, David E Rumelhart, Group et al. 1986) che prevede un'integrazione top-down ai dati elaborati in prospettiva bottom-up, con tre livelli di rappresentazione: fonetica, fonologia e semantica. È un modello localista in quanto prevede che esista una istanza mentale per ogni unità di rappresentazione e postula che venga a formarsi una traccia di una certa durata temporale che viene, poi, modificata con le informazioni in ingresso e confrontata con le rappresentazioni.

2.5.1 La plausibilità biologica: neuroni e reti neurali artificiali

Rivendicando un'architettura computazionale direttamente ispirata ad alcuni principi di base del cervello, il connessionismo capovolge in modo radicale la maggior parte delle assunzioni del paradigma cognitivista, il quale nei decenni precedenti ha (pre-)definito i criteri di studio dell'intelligenza artificiale in base al binomio regola-rappresentazione.

«Quel che piuttosto abbiamo trovato è che le informazioni sul *processing* in stile cerebrale si sono di per sé dimostrate molto stimolanti, nei nostri sforzi di costruire modelli. Ed è così che, in tutto il nostro lavoro, abbiamo posto al centro non la creazione di modelli neurali (e cioè, modelli di neuroni), ma piuttosto di modelli ispirati neuralmente dei processi cognitivi» (James L McClelland, David E Rumelhart, Group et al. 1986). Lo studio degli aspetti prettamente biologici è stato invece ampiamente condotto nell'ambito delle neuroscienze computazionali che utilizza modelli matematici e le reti neurali artificiali per poter comprendere ulteriormente principi e meccanismi alla base del funzionamento del sistema nervoso centrale²¹, di cui i neuroni rappresentano le unità funzionali. Illustrandone schematicamente la struttura, i neuroni sono composti da:

- un soma, ossia il corpo cellulare del neurone completo di nucleo;
- dendriti, che costituiscono la principale via di comunicazione in ingresso. Multipli per ogni neurone, ricevono e trasmettono le informazioni attraverso le variazioni di potenziale d'azione: se il potenziale di azione in ingresso supera un certo valore, generano a loro volta degli impulsi in grado di propagarsi attraverso l'assone del neurone stesso;

²¹Alcuni contributi importanti sono già contenuti in Parallel Distributed Processing e sono orientati allo studio della fisiologia della corteccia cerebrale e della plausibilità biologica degli algoritmi realizzati.

- l'assone costituisce la via di comunicazione che attraversa il neurone: attraverso la guaina mielinica e i nodi di Ranvier genera una trasmissione saltatoria;
- le sinapsi sono, infine, i punti di raccordo che permettono la trasmissione di segnali elettrici dal neurone corrente (o neurone pre-sinaptico) ad un altro neurone (o neurone post-sinaptico). A seconda dell'azione esercitata dai neurotrasmettitori (ammine, peptidi, aminoacidi, responsabili della modulazione degli impulsi nervosi), hanno una funzione eccitatoria oppure inibitoria.

I neuroni artificiali, o nodi della rete artificiale, riprendono e semplificano alcune caratteristiche tipiche dei neuroni biologici, senza tuttavia riuscire ad imitarne il funzionamento in maniera esaustiva, poiché ai fenomeni biochimici sono sostituite funzioni matematiche (tipicamente distribuzioni). Quando si parla di modellazione neuro-computazionale ci si riferisce alla modellazione dell'elaborazione e di memorizzazione mediante reti neurali artificiali di ispirazione biologica, per cui diventa sostanziale il concetto di APPRENDIMENTO che si distingue principalmente in due diversi modelli a seconda della modalità di presentazione degli esempi: (i) l'apprendimento supervisionato, (ii) l'apprendimento non supervisionato. Il diverso algoritmo di apprendimento presuppone dirette ripercussioni sull'architettura interna del modello computazionale.

Volendo fornire una veloce distinzione tra i due tipi di apprendimento, una rete neurale è detta supervisionata quando impara a fornire le risposte appropriate per ciascuno stimolo d'ingresso, modificando i valori delle proprie connessioni sinaptiche in base all'apprendimento di caratteristiche invarianti degli stimoli di ingresso James L McClelland, David E Rumelhart, Group et al. 1986. Una rete neurale non-supervisionata, che apprende sostanzialmente per auto-organizzazione, è invece in grado di estrarre informazione dall'input stesso, classificandolo autonomamente secondo pattern ricorrenti, sviluppando configurazioni interne simili a quelle dei circuiti neurali, come risposta adattiva all'insieme dei dati di input. Affinché sia efficace questo genere di modellazione, è necessario che vi sottendano alcuni requisiti minimi: la definizione dell'input, l'indipendenza dell'algoritmo dalla struttura della lingua in esame

e/o dalla codifica, e la definizione del processo di addestramento della rete²². In un approccio SUPERVISIONATO, solitamente si distinguono tre tipi di rappresentazioni: i nodi al livello di input, che ricevono direttamente dall'esterno i dati del problema da risolvere o del fenomeno da modellare; i nodi al livello di output, che forniscono la soluzione del problema; i nodi del livello interno (tipicamente nascosti) interposti tra quelli di input e quelli di output. L'apprendimento, in questo caso, avviene attraverso l'algoritmo di retropropagazione dell'errore (Error Back-Propagation), generalmente usato in combinazione con un metodo di ottimizzazione, quale la discesa stocastica del gradiente²³: la rete riceve un campione di coppie come pattern di input e di output, come soluzione della corrispondente combinazione di ingressi, ispirandosi ai metodi statistici di regressione. Nello specifico, vengono adattati i pesi delle connessioni fra nodi finché non si raggiunge il valore di uscita atteso; l'apprendimento della rete neurale, ossia la riduzione della distanza tra i valori d'uscita e i valori desiderati, consiste quindi in questa ricerca dei valori minimi della funzione d'errore. L'algoritmo, dunque, richiede un'uscita desiderata per ogni valore in ingresso. L'algoritmo di Back-Propagation ha evidenziato un sensibile incremento delle capacità di apprendimento delle reti neurali artificiali, ma non è privo di criticità; sebbene l'apprendimento su reti multistrato rappresenti uno straordinario strumento matematico-statistico di approssimazione e ottimizzazione in sistemi non lineari, evidenzia il problema della definizione di output atteso insieme all'input.

Viceversa, un modello computazionale che si ispira alla "legge di Hebb" (1949) prevede una modulazione adattiva della risposta sinaptica, implementando la capacità adattiva delle unità neuronali in contesti di apprendimento associativo, in modo NON-SUPERVISIONATO. La formulazione originale di Hebb postula – da un punto di vista meramente teorico – che se uno degli assoni terminali della cellula A (presinaptica) prende parte ripetitivamente nel processo

²²Tipicamente viene definita innanzitutto la rappresentazione interna dell'input, poi il numero di epoche di addestramento ed infine il decadimento della plasticità.

²³È un metodo iterativo per l'ottimizzazione di funzioni differenziabili, approssimazione stocastica del metodo di discesa del gradiente (GD) quando la funzione costo ha la forma di una somma. SGD opera similmente a GD ma, ad ogni iterazione, sostituisce il valore esatto del gradiente della funzione costo con una stima ottenuta valutando il gradiente solo su un sottoinsieme degli addendi. La discesa del gradiente classica viene anche chiamata batch GD, in quanto l'intero dataset viene valutato ad ogni iterazione. Quando il gradiente viene invece approssimato stocasticamente, è possibile costruire diversi metodi a seconda del numero di addendi usati ad ogni iterazione.

di eccitamento della cellula B (postsinaptica), qualche cambiamento strutturale o metabolico subentra in modo che l'efficienza di A, come cellula eccitatrice di B, aumenti. Un ulteriore contributo di Hebb alle neuro scienze computazionali è stata quella di espandere la sua regola a gruppi di neuroni: eccitandosi insieme questi formano raggruppamenti cellulari associativi.

Per gli scopi del presente lavoro, l'ipotesi di apprendimento adattivo basato sulla plasticità risulta particolarmente importante; e tale apprendimento adattivo è tipicamente alla base dell'apprendimento NON-SUPERVISIONATO.

2.5.2 Apprendimento non-supervisionato: mappe auto-organizzanti

Una rete neurale non-supervisionata è in grado di estrarre informazione dall'input in assenza di informazioni classificatorie; pertanto una rete addestrata con algoritmo non-supervisionato è in grado di classificare autonomamente pattern ricorrenti presenti nel dato di input, sviluppando configurazioni interne simili quelle dei circuiti neurali, come risposta adattiva all'insieme dei dati di input. Tra i modelli non supervisionati, le reti auto-organizzanti sono tipicamente adattive e in grado di catturare la correlazione incrementale tra strutture ricorrenti dell'input, risposte abituali della rete in attivazione, e una fase di radicamento delle risposte in catene di attivazione memorizzate a lungo termine. In una tale implementazione, la struttura morfologica può emergere durante l'apprendimento proprio a partire dai principi di auto-organizzazione dell'input, in funzione del supporto probabilistico che ogni forma riceve dall'insieme delle forme ad essa morfologicamente collegate (famiglia di vicini morfologici)²⁴. Ad assumere una certa rilevanza sono i fattori distribuzionali della *token* e della *type frequency*, oltre la coerenza formale dei paradigmi flessionali e delle famiglie lessicali che modulano la competizione — in termini di agevolazione o inibizione nell'attivazione di sequenze simboliche nella memoria a breve termine (o *Working Memory*, WM) — con lo scopo di integrare queste sequenze selettivamente con le aspettative della memoria a lungo termine (o *Long Term Memory*, LTM). L'idea di un lessico inteso come sistema complesso adattivo può essere implementata con le *Self-Organising Maps* (SOMs, Kohonen 2001 su apprendimento auto-organizzato, dove è somministrato solo l'input, e sono definite delle regole di plasticità. Sempre relativamen-

²⁴La ridondanza delle strutture in questo caso condiziona la percezione della rete in termini di predizione/aspettativa (in modo del tutto simile a quanto l'evidenza comportamentale dei parlanti suggerisce).

te all'idea della plausibilità biologica, la rete di Kohonen prende le mosse dalla capacità del cervello di comprimere e organizzare dati sensoriali in maniera spontanea e alla tendenza dei neuroni ad eccitare i neuroni adiacenti e ad inibire quelli meno vicini (Zeidenberg 1990). I modelli ispirati a questo principio sono ritenuti biologicamente plausibili, perché di fatto non hanno bisogno di informazioni esterne circa la classificazione, e non apprendono in base al principio della Back-Propagation. Quella di Kohonen è una rete neurale dall'architettura sia feed-forward che feedback e la sua caratteristica peculiare è la capacità di modificare la configurazione dei propri nodi in base al peso che assumono man mano che vengono forniti gli input: i nodi con pesi simili si avvicinano, quelli con pesi molto diversi si allontanano. In definitiva una SOM emula il comportamento dei cluster di neuroni delle aree corticali coinvolti nella elaborazione e classificazione di stimoli: la fase di elaborazione consiste nell'attivazione dei neuroni (e quindi nodi) in risposta a stimoli diversi e l'auto-organizzazione è orientata alla progressiva e dinamica specializzazione di nodi diversi che si attivano a stimoli diversi. La co-attivazione di nodi vicini, in risposta a stimoli simili, avviene per i nodi sul layer di output: nell'addestramento per apprendimento competitivo, quando è somministrato uno stimolo alla rete si calcola la sua distanza euclidea da tutti i vettori dei pesi; il nodo col vettore dei pesi più simile all'input è il nodo vincente, definito Best Matching Unit (o BMU), ed è quello con il livello di attivazione maggiore, mentre i pesi dei nodi vicini decrescono nel tempo e in funzione della distanza dalla BMU (definita da una funzione gaussiana).

Una variante della SOM classica arricchita di un livello di connessioni rientranti definite su un layer temporale, è noto come mappa auto-organizzante temporale, o Temporal Self-Organising Map (TSOMs, Ferro, Marzi e Pirrelli 2011; Marzi, Ferro e Pirrelli 2014; Pirrelli, Ferro e Marzi 2015). Caratteristica peculiare della TSOM è che mantiene le informazioni seriali nel tempo attraverso catene di attivazione predittiva di recettori che codificano le informazioni spaziali e soprattutto temporali degli stimoli di input²⁵. Attraverso connessioni ricorrenti, la TSOM ricodifica gli stimoli di ingresso in modo auto-associativo, sfruttando la ridondanza formale delle SERIE TEMPORALI di simboli. L'apprendimento, e quindi la memorizzazione, con-

²⁵Sono previsti due livelli di connettività: spaziale (*what*), dal livello di input a ogni nodo della mappa e temporale (*when*), rientrante da ogni nodo a tutti gli altri della mappa. Le parole sono rappresentate come serie temporali di simboli, somministrati un simbolo alla volta.

siste dunque nella trasformazione adattiva di una risposta occasionale in una rappresentazione interna di quella parola, che si attiva sistematicamente in risposta all'input di quella parola. L'informazione espressamente ridondante e basata sul contesto massimizza il numero di *Best Matching Units* affinando le capacità predittive della mappa. Più tracce di memoria simultaneamente attive definiscono una dinamica di competizione, per cui se una rappresentazione parziale (intesa come sequenza temporale di BMU) è evocata sistematicamente da un input, non significa necessariamente che si attivi in risposta a quel solo input, così come uno stesso stimolo può attivare simultaneamente più di una rappresentazione nella rete; a questo punto il livello di co-attivazione offre una misura del grado di similarità percepita tra rappresentazioni interne. Entrando nel dettaglio degli scopi di questo lavoro, a parole che condividono la stessa struttura morfologica possono corrispondere rappresentazioni interne fortemente e sistematicamente co-attivate. Ciò porta alla rivalutazione del ruolo della struttura morfologica nel lessico, portando a riconoscere i morfemi come strutture ricorrenti di catene temporali di BMU e l'acquisizione lessicale come processo dinamico basato sull'individuazione emergente di quelle strutture formali condivise dalle famiglie. Da questo punto di vista, l'apprendimento discriminativo si rivela una strategia efficace per strutturare il flusso di input in rappresentazioni interne finalizzate al riconoscimento e alla produzione delle parole. Il vantaggio di questi modelli è che offrono un correlato funzionale dei meccanismi di attivazione di stimoli lessicali concorrenti (Joan L. Bybee 1995, James P. Blevins 2016, Burzio 2004) e soprattutto si accordano con le basi neuro-funzionali dell'interazione emergente tra percezione, memoria di lavoro e memoria a lungo termine.

2.6 Problemi teorici della derivazione

Ho già ampiamente dibattuto nel corso di questo capitolo sulla convergenza dei recenti approcci teorico-sperimentali a favore di una visione della competenza morfologica riguardante i meccanismi che stanno alla base della percezione delle relazioni sintagmatiche e paradigmatiche delle parole morfologicamente complesse, tra cui anche le parole derivate.

Ovviamente, la morfologia derivativa è un campo vasto e l'elenco dei processi o delle relazioni formali pone ancora molte sfide, sia empiriche che teoriche, ma sulla base delle evidenze

psicolinguistiche ottenute negli ultimi anni (Lieber 2017; Burani e Anna M Thornton 2011) è possibile focalizzare l'attenzione sui processi di suffissazione per scorgere alcune caratteristiche nei processi di formazione e percezione delle parole. Nel merito delle recenti indagini e della presente tesi, proprio la suffissazione è presa in esame per determinare se anche i suffissi e gli schemi morfologici implicati in derivazione svolgano un ruolo nell'accesso e nell'elaborazione di parole complesse, e se eventualmente la salienza del suffisso e la grandezza della serie influenzino la loro percezione e elaborazione. A questo proposito bisogna tenere conto di quelle che sono le caratteristiche formali e distribuzionali dei suffissi, come ad esempio il grado di trasparenza; se da un lato la flessione ha una forte tendenza ad esibire una piena composizionalità semantica nel senso fregeano²⁶ del termine, la derivazione mostra spesso problematiche di opacità, motivo per il quale possibilmente esiste una differenza nel processo di percezione della parola. Secondo la Hay (2001) la segmentabilità morfologica è anche influenzata dalle proprietà frequenziali delle parole stesse e delle sue unità costituenti, che a loro volta spesso vanno di pari passo con gradiente fonologico e opacità semantica²⁷, intendendola come fenomeno di gradiente (Gonnerman, Mark S Seidenberg e Andersen 2007). Mentre gli approcci teorici alla morfologia derivazionale sono stati in gran parte interessati alla forma, alla struttura, al significato e alla relazione tra loro, gli approcci psicolinguistici – come già più volte menzionato - si sono concentrati su PERCEZIONE, ELABORAZIONE e PRODUZIONE. Una domanda che è stata centrale e che si ripropone nello studio psicolinguistico della derivazione è se le parole siano scomposte nei loro morfemi costitutivi quando vi si accede nel lessico mentale o se si acceda alle parole derivate come forme intere, non analizzate come “scomponibili”. I primi lavori sperimentali (Taft e Forster 1975) in questo senso suggeriscono che le parole vengono analizzate nei loro morfemi componenti durante l'accesso lessicale, almeno per le parole formate con prefissi. Il *Parallel Dual Route* (Schreuder e R. H. Baayen 1995) propone che i due tipi di accesso lessicale (decomposizionale e parola intera) non sono mutuamente esclusivi, ma

²⁶Secondo Frege per capire il riferimento di un intero enunciato è necessario rivolgere l'attenzione alle parti che lo compongono. Il principio di composizionalità è uno dei capisaldi della teoria fregeana del significato (Frege 1948).

²⁷Il principio della frequenza relativa (Hay 2001) riguarda l'ipotesi della frequenza assoluta, per cui parole derivate ad alta frequenza tendono ad essere meno trasparenti (più irregolari) nel loro significato e nella loro struttura.

piuttosto operano allo stesso tempo e competono tra loro: le parole complesse possono essere memorizzate come intere o elaborate nelle loro parti, con una serie di fattori (frequenza della parola derivata, frequenza della base, dimensione della famiglia, ecc.) che determinano quale tra il percorso a parola intera o il percorso di decomposizione, risulta come quello “vincente” (Lieber 2017).

2.6.1 Effetti di frequenza

La letteratura psicolinguistica ha identificato le caratteristiche di natura distribuzionale-formale che influenzano maggiormente la codifica di una parola. Tra le caratteristiche più influenti, gli effetti di frequenza sembrano essere determinanti e sono stati trovati soprattutto con compiti di decisione lessicale: secondo la legge di Zipf²⁸ le parole più frequenti sono anche le più corte, inoltre un numero estremamente limitato di parole, essenzialmente grammaticali, ha frequenza molto alta, mentre sono numerose le parole lessicali che tendono a ricorrere con una frequenza molto più bassa²⁹. Si stima che rispetto alle parole a bassa frequenza quelle ad alta frequenza vengono comprese più rapidamente (con differenze dell'ordine di 100-150 ms nei compiti di decisione lessicale), fissate più brevemente ed evocate in una finestra temporale che si aggira sui 132 ms dopo l'inizio del target; tuttavia, a seconda dei modelli di riferimento sussistono evidenze diverse. Ciò che è importante stabilire a priori, però, è che l'effetto di frequenza non tocca solo la parola come unità, ma anche basi e affissi. Ad esempio, a parità di frequenza complessiva di parole, quelle con una base più frequente sono classificate in tempo minore come parole (nel tipico esperimento di *lexical decision* utilizzato nei protocolli sperimentali con soggetti), rispetto a quelle con una base meno frequente.

²⁸Nel 1935 il linguista statunitense George Kingsley Zipf notò un curioso schema nella frequenza delle parole. In un qualsiasi testo la parola più frequente appare due volte più spesso della seconda più frequente, e la terza parola più frequente frequenza appare un terzo delle volte rispetto alla prima, la quarta un quarto e così via. La legge di Zipf formalizza – come segue – il legame matematico che sussiste tra la frequenza ed il rango di una parola: $f(z) = C/za$ dove $f(z)$ è la frequenza della parola di rango z , C è la costante ed a è l'indice inverso di ricchezza lessicale, vale a dire che più grande è a , più ristretto risulta il vocabolario del corpus.

²⁹Sono definite hapax legomenon (lett. «detto una sola volta», gr. hápax legómenon) le parole che occorrono una sola volta.

2.6.2 Numerosità e produttività

Considerando la frequenza dei suffissi, si possono adottare due principali misure quantitative. Da un lato, la frequenza in senso proprio viene calcolata sui *word-token*, sommando la frequenza cumulata - in un dato corpus- delle occorrenze delle parole con un dato suffisso. D'altra parte, la frequenza del suffisso può essere misurata calcolando il numero di parole tipo (*word-type*) in cui si verifica un dato suffisso in una data lingua. Questa seconda misura determina la numerosità del suffisso, ovvero il numero totale di parole diverse con quel determinato suffisso esistenti nella lingua. La numerosità sembra essere una caratteristica quantitativa importante oltre che predittiva delle modalità di accesso, poiché un suffisso “numeroso”, che compare in molte combinazioni diverse con basi diverse, potrebbe favorire l'emergere del suffisso come unità di elaborazione indipendente. La numerosità è inoltre strettamente correlata alla produttività del suffisso stesso (H. Baayen 1992), ovvero la possibilità per un qualsiasi procedimento morfologico in una data lingua di essere utilizzato per la costituzione di nuove formazioni (può trattarsi sia di lessemi sia di forme flesse) della lingua in questione. La produttività è una proprietà graduale, e non polare: è possibile, dunque, l'esistenza di procedimenti morfologici con gradi diversi di produttività³⁰. Basandosi sulle distribuzioni zipfiane, Baayen (1992), in particolare, ha osservato che i procedimenti morfologici più produttivi sono rappresentati, nei corpora linguistici, da un ampio numero di forme a scarsa frequenza e da un numero piuttosto piccolo di forme con frequenza alta; i procedimenti poco produttivi sono caratterizzati da una distribuzione inversa, cioè molte forme ad alta frequenza e poche forme a bassa frequenza. È possibile fornire una motivazione per questa distribuzione in termini psicolinguistici, a partire dai modelli del lessico mentale dapprima descritti.

Immaginando che l'accesso a una parola morfologicamente complessa nel lessico mentale possa avvenire tramite due modalità o “percorsi” (Cfr § 2.2), si presuppone dunque che il lessico mentale contenga parole intere, ma anche basi lessicali e morfemi legati (ad esempio

³⁰L'applicazione di qualsiasi procedimento morfologico è soggetta sia a restrizioni interne al sistema linguistico (strutturali), sia di tipo extralinguistico. Questo insieme di fattori viene trattato diversamente a seconda dell'impostazione teorica adottata: può essere compreso totalmente o parzialmente nella definizione generale di produttività di un procedimento, oppure può esserne escluso e considerato come attinente a un livello diverso rispetto a quello dove si realizza la produttività (Dressler e Ladányi 2000).

suffissi e prefissi). A seconda del modello, le due modalità possono essere in competizione diretta tra loro oppure contribuire in parallelo al recupero della forma, ma in tutti i casi resta valido il presupposto che, per ogni occasione di recupero lessicale, un percorso sarà dominante rispetto all'altro. Ogni volta che avviene l'accesso ad una forma, questo ha come conseguenza un rafforzamento della sua rappresentazione nel lessico mentale. Questo rafforzamento in genere viene rappresentato come un innalzamento del livello di attivazione della rappresentazione neurale della forma stessa. Se dunque una forma morfologicamente complessa è molto frequente, cioè l'accesso ad essa avviene molto spesso, la sua rappresentazione come forma intera si rafforza nel lessico mentale. Di conseguenza, il percorso di accesso diretto a quella forma diventa predominante durante l'elaborazione linguistica: in pratica, per accedere a quella forma non è (più) necessario che venga scomposta in morfemi. Questo favorisce, col tempo, una opacizzazione della forma complessa sia dal punto di vista morfo-fonologico sia dal punto di vista morfo-semantic, fino a che (nel caso estremo) non viene più percepita come complessa. Al contrario, una forma poco frequente difficilmente ha una rappresentazione autonoma ben radicata nel lessico mentale, per cui l'accesso potrebbe avvenire tramite scomposizione morfologica. In questo modo sono i morfemi che la compongono, e in particolare eventuali affissi, ad essere attivati e dunque rafforzati nella mente del parlante, il che contribuisce alla loro produttività. Infatti, questo rafforzamento dà loro un vantaggio che entra in gioco durante l'elaborazione di qualsiasi forma complessa costituita con quei morfemi, e rende più probabile in tutti quei casi l'accesso tramite scomposizione. Alcuni esperimenti psicolinguistici (Cfr. Gaeta e Ricca 2003) hanno posto il focus sulle cause della produttività dei procedimenti morfologici (in particolare di quelli derivazionali) nel modo in cui il cervello elabora l'input linguistico, trovando una forte correlazione tra frequenza e produttività. Tuttavia, Baayen stesso (1992) ha notato che una simile spiegazione, benché sostanzialmente adeguata, non rende pienamente conto dei dati psicolinguistici che la ricerca ha a disposizione. Ad esempio, Burani e Caramazza (1987) documentano che effettivamente la frequenza delle parole complesse ha una correlazione positiva con la rapidità del loro riconoscimento, ma che l'altro fattore determinante è la frequenza delle basi ricomprese nelle parole complesse. Hay (2001) argomenta che ad influenzare la trasparenza (e dunque la produttività) di una forma complessa non è la sua frequenza assoluta, bensì la sua frequenza relativa in confronto a quella della

base. Se abbiamo a che fare con una forma complessa e la sua base è più frequente della forma complessa intera, indipendentemente dalla frequenza assoluta di entrambe, tra le due sarà la rappresentazione della base ad avere un livello di attivazione più alto, e dunque l'accesso tramite scomposizione sarà favorito rispetto ad un accesso diretto. Al contrario, se la parola complessa è più frequente della sua base, anche se entrambe hanno una frequenza assoluta nel range delle basse frequenze, sarà comunque l'accesso diretto ad avere un vantaggio, perché il livello di attivazione della forma intera rimane più alto di quello della base.

Capitolo 3

METODI E DATI

Il presente lavoro è orientato alla modellazione del livello di organizzazione morfologica emergente nel lessico mentale. Specificamente è preso in esame il compito di percezione e identificazione dei componenti morfologici all'interno di forme – per l'appunto – morfologicamente complesse, in funzione delle interazioni tra rappresentazioni, fattori distribuzionali del lessico e condivisione della struttura nei dati lessicali. Le dinamiche di memorizzazione ed elaborazione, nei processi di ricodifica degli input e nelle strategie adattive per l'organizzazione della memoria a lungo termine, sono indagate in un quadro computazionale basato su mappe auto-organizzanti con connessioni hebbiane, definite su uno strato temporale (i.e. mediante *Temporal Self-Organising Maps*, o TSOM).

Le simulazioni computazionali offrono, infatti, la possibilità di investigare e modellare la stretta relazione tra strategie di memorizzazione (*storage*) e di elaborazione (*processing*), in condizioni di input controllate, anche sulla base di modelli teorico-descrittivi.

In questo capitolo verrà delucidato, a seguito di una breve disamina sui corpora di riferimento e sui criteri di selezione dei dati, in che modo un'auto-organizzazione fortemente adattiva, in fase di addestramento, possa determinare se gli schemi morfologici, per determinati suffissi, svolgono un ruolo nell'accesso e nell'elaborazione di parole complesse italiane. A queste descrizioni si accodano interessanti interrogativi riguardanti il ruolo di variabili quali la salienza del suffisso e la grandezza della serie rispetto all'elaborazione, la capacità di ricodificare e mantenere serie temporali di unità simboliche (i morfemi) nella memoria di lavoro, e infine l'organizzazione di queste rappresentazioni nella memoria a lungo termine.

3.1 Risorse e strumenti: i corpora

I corpora sono grandi collezioni di testi che rappresentano l'uso naturale della lingua¹. Alcuni specifici parametri definiscono le diverse tipologie di corpora per permettere innumerevoli analisi in base ai casi di studio. Esistono corpora specialistici che contengono testi appartenenti a un certo dominio che solitamente vengono utilizzati in un ristretto campo di ricerca e ci sono per contro corpora generali che sono definiti plurifunzionali ma soprattutto risorse trasversali di riferimento, proprio perché l'alto livello di generalità permette uno studio comprensivo della lingua nel suo complesso. Altre distinzioni circa i corpora riguardano la modalità, la lingua, la cronologia, l'integrità e la codifica dei testi. Con un approccio prettamente statistico-acquisizionale, il corpus viene definito campione (o *sample*) rappresentativo della lingua o di una varietà, per poter studiare e generalizzare induttivamente fenomeni linguistici di interesse, dal campione alla popolazione. Per garantire la rappresentatività del corpus è necessario che questo sia bilanciato e perché ci sia il bilanciamento è necessario che siano presenti le descrizioni riguardanti i confini spaziali e temporali dei testi e la loro tipologia. Il concetto di corpus come campione è essenziale, giacché sia la dimensione quantitativa che quella qualitativa incidono fortemente sul modo in cui i dati ricavati possono essere proiettati in un modello della lingua (Lenci, Montemagni e Pirrelli 2005).

3.1.1 Il DerIvaTario e il CoLFIS

Per gli scopi specifici di questo studio, i due corpora di riferimento sono il DerIvaTario (Talamo, Celata e P. M. Bertinetto 2016) e la fonte lessicale dalla quale attinge, il CoLFIS, *Corpus e Lessico di Frequenza dell'Italiano Scritto* (P. M. Bertinetto et al. 2005). Questo è un corpus lemmatizzato di tre milioni di occorrenze (*tokens*) di italiano scritto, campionato da una varietà accuratamente bilanciata di testi, riviste e giornali.

¹Attualmente con il termine corpus si implica il suo formato digitale, ma la creazione di un corpus come raccolta di testi è un'operazione ben più attempata per gli studi sul linguaggio. L'avvento del computer ha permesso di collezionare dati linguistici sempre in maggior quantità, di facilitarne lo studio e di stilare modelli computazionali di lingua. Gli albori di questo approccio si riconoscono a Padre Roberto Busa, con un'analisi lessicografica dell'opera omnia di Tommaso d'Aquino.

CoLFIS nasce prettamente con lo scopo di rappresentare il lessico mentale del parlante italiano nella maniera più verosimile possibile (Alessandro Laudanna, Anna Maria Thornton et al. 1995), distinguendosi da molti dizionari di frequenza dell'italiano scritto in uso (es. Bortolini, Tagliavini e Zampolli 1971), sia per la quantità di lemmi (circa quattro milioni contro le 500.000 parole del LIF) che per il bilanciamento delle fonti.

“Quando in un contesto psicolinguistico si parla della frequenza di parola, ci si riferisce ad una ideale frequenza media, che faccia astrazione da particolari idiosincrasie, frequenze soggettive, sovraesposizioni di talune categorie di persone a lessici specialistici, etc. È perciò necessario che un lessico di frequenza centrato sul ricevente non sia basato su tipi testuali stabiliti a priori o campionati in maniera distorta, ma su criteri di selezione del corpus che permettano di far emergere per quanto possibile la frequenza media di ricezione.” (Alessandro Laudanna, Anna Maria Thornton et al. 1995)

La costituzione di DerIvaTario a partire da questo Formario, si presenta come un'estensione morfologica di questo, condividendone il lessico, le informazioni sulla frequenza e sulla classe di parola, ma concentrandosi prettamente sui meccanismi di tipo affissionale. Il derIvaTario è infatti organizzato come uno schema di annotazione per la morfologia derivazionale con informazioni morfologiche, morfotattiche e morfosemantiche riguardanti la base del derivato². Questo database interattivo è miratamente adeguato agli studi sulla morfologia teorica e nella fattispecie per la ricerca psicolinguistica (Talamo e Celata 2011) ed è particolarmente funzionale nel tenere conto di diverse proprietà morfologiche della base e di ciascun affisso coinvolto nei processi derivazionali³. Pur proponendosi come lessico per l'indagine psicolinguistica, derIvaTario non contiene informazioni sull'elaborazione dei derivati e non pone una differenza tra quelli immagazzinati nel lessico mentale e quelli elaborati in tempo reale. Riguardo l'annotazione di tipo morfologico, invece, il discorso verte sull'identificazione degli affissi, dei loro allomorfi e sulla descrizione dell'allomorfia della base; ogni affisso è identificato in base a criteri morfosemantici ed è corredato dal morfema che lo identifica nell'annotazione (di fatto, corrisponde all'allomorfo più utilizzato).

²Questo schema è stato impiegato nell'annotazione di circa 11.000 derivati italiani estratti dal CoLFIS e l'accordo tra annotatori è stato valutato su diverse variabili dello schema di annotazione.

³In modo cruciale la trasparenza morfotattica e morfosemantica (per quest'ultima si veda Gary Libben 1998).

3.2 Il dataset di addestramento: i criteri di selezione

La definizione del tipo di dati estratti si basa sui principali interrogativi relativi all'aspetto percettivo che si evidenziano nelle operazioni di derivazione in *-zione*, *-tore* e *-ico*. Questi tre suffissi, di diversa salienza, presentano un'alternanza di complesse configurazioni strutturali che evidenziano, da un punto di vista percettivo, le difformità emergenti nella loro natura di affissi (più o meno fusi alla base). Sebbene il primo esperimento sia un addestramento senza le frequenze da corpus (al fine di evidenziare effetti prettamente strutturali), si è optato per una condizione a frequenze bilanciate per somministrare alla rete una token frequency uniforme pari a 5 e limitare possibili effetti di *primacy* e *recency*, altrimenti evidenti in totale assenza di frequenze. Il secondo addestramento è specificamente pensato con frequenze atte a evidenziare quanto la *token frequency* possa bilanciare o eventualmente accentuare certi effetti. Nella scelta dei suffissi si è tenuto conto di una possibile sovrapposizione di famiglie, oltre al grado di produttività del morfema (dove per produttività si intende il numero di basi cui si può applicare un determinato morfema in una lingua), laddove le serie presentano un diverso grado di consistenza funzionale. Questa scelta è stata operata nella eventuale possibilità di esaminare anche gli effetti delle grandezze di famiglia e serie (esperimento 3). Sono state selezionate, per ciascuno dei suffissi menzionati con le relative allomorfie, 200 forme di cui è stata riportata la frequenza⁴ riferita al lemma (pertanto sono incluse eventuali forme diversamente flesse, sia nel *derIvaTario* che nel *CoLFIS*), l'allomorfo⁵ per calcolare il punto di confine di morfema, la base lessicale e l'elemento n-esimo del vicinato, anch'essi corredati di frequenza estratta rispettivamente da *DerIvaTario* e *CoLFIS*⁶. Nello studio del fenomeno percettivo della derivazione di questi suffissi, il tentativo è di comprendere se vi sia una diversa percezione della struttura interna delle parole derivate, sia in base alla diversa distribuzione dell'input, sia in base alla diversa salienza dei suffissi derivazionali in esame.

⁴Mirando a un addestramento con frequenze, rispetto alla grandezza delle famiglie, sono stati evitati gli hapax.

⁵A tale proposito, l'allomorfo più utilizzato sarà quello identificato come l'etichetta della forma. Ad esempio il processo derivazionale in *-(z)ione* conta tre allomorfi: *-ione*, *-gione* e *-zione*. Tuttavia l'allomorfo più utilizzato è *-zione*, per cui questo processo derivazionale viene in ogni caso identificato nell'etichetta *-zione*.

⁶La base, spesso individuata nel verbo, non è stata considerata come membro della famiglia o vicinato, per ovviare a criticità di altra natura legate all'infinito inteso come lemma del paradigma oppure come forma flessa.

3.2.1 Criteri di annotazione

L'input che la rete riceve è costituito dalle forme, unitamente alla loro frequenza (bilanciata pari a 5, o secondo la distribuzione nel corpus di riferimento). Non vi è alcuna informazione circa la struttura morfologica. Al solo fine delle analisi quantitative riportate nel capitolo successivo (Capitolo 4), sono annotate una serie di informazioni funzionali alle valutazioni successive all'addestramento.

Nel dettaglio, si considerano per l'addestramento:

- l'input somministrato un simbolo alla volta, trattandosi di una rete ricorrente, con simbolo di inizio e fine parola, come ad esempio #,A,B,I,T,A,Z,I,O,N,E,\$ per *abitazione*;
- la *token frequency*, ovvero la frequenza con la quale ogni singola parola sarà somministrata alla rete per ognuna delle epoche di addestramento.

Per le successive analisi del comportamento della rete a fine addestramento, ho considerato:

- il suffisso derivazionale, ovvero, nel caso specifico, se la forma appartiene alla serie derivazionale in *-zione, -tore, -ico*;
- la base lessicale, ovvero l'esponente lessicale del processo derivazionale;
- la lunghezza della base, ovvero della forma senza il suffisso derivazionale;
- la lunghezza del suffisso;
- l'eventuale allomorfo impiegato nel processo derivazionale, al posto del suffisso derivazionale più frequente;
- la famiglia (per il solo Esperimento 3) e la sua (ir)regolarità.

3.3 La Temporal Self-Organising Map

Le mappe auto-organizzanti sono strumenti ideali per la rappresentazione dell'organizzazione morfologica, e possono rispondere alla domanda se sia possibile per un parlante pervenire a

un'organizzazione "globale" dei diversi fenomeni morfologici della sua lingua (sfruttando anche le similarità tra le parole memorizzate), sulla base di una strategia di elaborazione lessicale "locale".

Le relazioni tra forme morfologicamente complesse possono essere rappresentate mediante strutture ad albero e a grafo, che consentono di modellare a livello algoritmico sia il radicamento di forme memorizzate singolarmente, sia le relazioni tra le forme memorizzate in simultanea, dove il tipo di rappresentazione formale è determinato dalla percezione della ridondanza morfologica. In una interpretazione probabilistica della struttura, la forza delle connessioni è risultato della frequenza di occorrenza di una specifica sequenza, sia nello stesso contesto che in contesti diversi. Quanto più una sequenza è attestata, tanto maggiore è la forza delle connessioni tra le unità che la compongono; ciò favorisce una specializzazione delle catene di nodi fortemente connessi per le forme ad alta frequenza e la cui forza va a definire il livello di radicamento di una forma nel lessico. In termini simulativi la modellazione algoritmica dell'accesso al lessico, oltre ad essere influenzata da meccanismi di competizione tra WM e LTM, interagisce con le dinamiche globali di memorizzazione ed elaborazione dipendenti da fattori distribuzionali ritenuti psicolinguisticamente rilevanti, come la frequenza, la regolarità formale o la famiglia lessicale. L'impatto di questo modello sui principi di organizzazione ed elaborazione morfologica è indagato attraverso una serie di simulazioni che evidenziano le dinamiche tra memoria a breve termine per quanto riguarda l'attivazione, memoria a lungo termine per quanto concerne l'apprendimento e l'organizzazione morfologica delle forme memorizzate. Una rete auto-organizzante, quindi adattiva, consente di catturare la correlazione incrementale tra le strutture ricorrenti dell'input (frequenza), le risposte della rete in fase di attivazione (competizione)⁷, e il radicamento delle risposte in catene di attivazione memorizzate a lungo termine (aspettative): la struttura morfologica emerge durante l'auto-organizzazione dell'input e in funzione delle relazioni probabilistiche tra una forma e quelle ad essa morfologicamente relate.

⁷La competizione tra nodi nell'addestramento per learning competitivo si risolve quando il nodo vincente (Best Matching Unit) è quello con il livello di attivazione maggiore. Somministrato un input alla rete, nel calcolo della sua distanza euclidea da tutti i vettori dei pesi, il neurone col vettore dei pesi più simile all'input sarà la BMU.

3.3.1 Architettura della rete

La TSOM è una rete ricorrente auto-organizzante costituita da una griglia bidimensionale di nodi organizzati topologicamente, che sono sia le unità di memoria che di elaborazione della rete, i quali non sono specificamente cablati per rispondere a determinati simboli. Dunque, una rete TSOM può essere addestrata con diversi stimoli. I nodi della rete, o neuroni artificiali, sono caratterizzati da due livelli di connettività: (i) tutti i nodi sono completamente connessi al vettore in input senza ritardo temporale (strato di connessioni spaziali); (ii) ogni nodo è connesso con tutti gli altri nodi (strato di connessioni temporali) (Ferro, Marzi e Pirrelli 2011; Pirrelli, Ferro e Marzi 2015; Marzi, Ferro, Cardillo et al. 2016; Marzi, Ferro e Pirrelli 2019). L'approccio fornisce un quadro generale in cui l'elaborazione e l'acquisizione lessicale sono implementate sia come strategie di ricodifica, che di memorizzazione per serie temporali di simboli, dipendenti da fattori specificamente linguistici oltre che da funzioni cognitive extra-linguistiche (organizzazione lessicale, accesso e recupero lessicale). Una TSOM consiste nell'organizzazione spaziale e temporale dei pesi su più livelli di connettività⁸; le connessioni spaziali ottengono informazioni sincrone dal livello di input, in cui ogni stimolo (quindi ogni simbolo) è codificato come un vettore. Le connessioni temporali simulano le sinapsi neurali con propagazione del ritardo e con pesi che determinano l'influenza che l'attivazione di un nodo al tempo t ha sull'attivazione dei nodi al tempo $t + 1$. In questo modo, le connessioni temporali trasmettono l'aspettativa probabilistica di attivazione di nodi specifici in base al corrente stato di attivazione della mappa (figura 3.1). Per ogni nodo sulla mappa, la sua risposta di elaborazione ad uno stimolo in ingresso al momento t è data da:

$$y_i(t) = \alpha \cdot y_{S,i}(t) + (1 - \alpha) \cdot y_{T,i}(t)$$

dove $y_{S,i}(t)$ è la quantità di attivazione dell' i -esimo nodo all'istante t che scorre attraverso le connessioni di ingresso, $y_{T,i}(t)$ è l'attivazione temporale dell' i -esimo nodo all'istante t innescata dallo stato di attivazione di tutti i nodi al tempo $t - 1$. Nell'equazione α e $(1 - \alpha)$ pesano il rispettivo contributo delle connessioni di ingresso (S) e delle connessioni temporali

⁸Si ricorda che le connessioni spaziali (definite anche come connessioni *what*) vanno dal livello di input a ogni nodo della mappa, mentre le connessioni temporali (definite come connessioni *when*) sono rientranti da ogni nodo a tutti gli altri della mappa.

(T) all'attivazione del nodo i . In modo più intuitivo, l'equazione integra lo stato di attivazione della mappa causato dal simbolo di input corrente con la quantità di aspettativa sollevata dall'esperienza del simbolo immediatamente precedente.

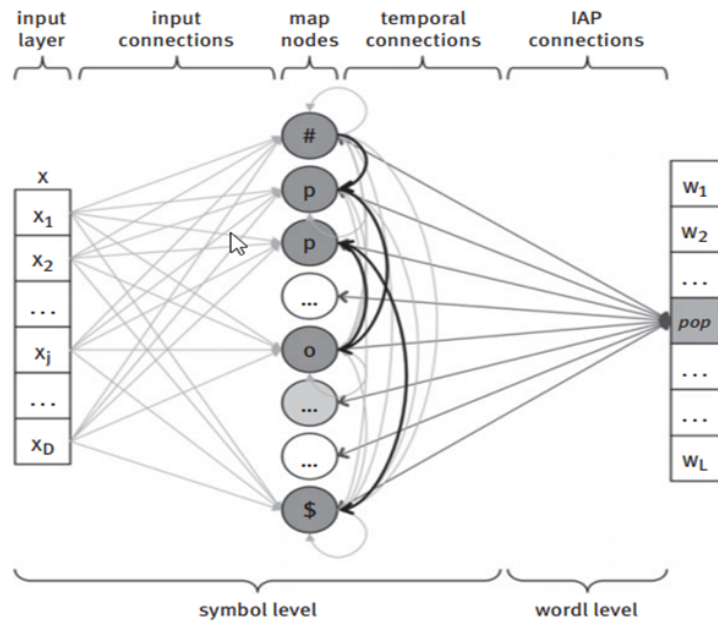


Figura 3.1:
Temporal Self-Organising Map sull'input della parola pop (“#pop\$”).
(tratto da Marzi, Ferro, Nahli et al. 2018)

3.3.2 Algoritmo di apprendimento

Il TRAINING consiste nel mostrare la mappa per ogni forma di input i simboli che la costituiscono una alla volta, ripetendo la somministrazione in funzione della distribuzione della forma stessa. Tipicamente, l'input è somministrato in ordine casuale (*bag of words*) per il numero di epoche, definite in maniera in parte empirica e in parte calcolata secondo la curva di decadimento del *learning rate*. Questa misura risulta funzionale all'errore, ma in qualche modo anche alla durata complessiva dell'addestramento (ovvero del numero totale delle epoche), definendo anche la curva di decadimento della plasticità della rete. Agli alti livelli iniziali della plasticità corrisponde un riaggiornino e adattamento continuo delle connessioni in funzione di ogni nuovo input; progressivamente la plasticità della rete diminuisce consentendole di sta-

bilizzare le connessioni e con esse le rappresentazioni interne alla rete in funzione dell'input. Il livello di input è implementato come un vettore binario, con uno schema di bit associato in modo univoco a ciascun simbolo (rappresentazione *one-hot*). Una parola viene presentata alla rete come sequenza seriale di simboli che inizia con un simbolo di inizio parola ('#') e termina con ('\$'). Lo stato di attivazione temporale della rete viene inizializzato alla presentazione di ogni nuova parola. Ciò implica che lo stato di attivazione della rete dopo aver visto la forma di parola di input corrente non ha memoria a breve termine delle parole appena passate. Somministrando l'input una volta soltanto, infatti, va da sé che le prime parole sono quelle che vengono cancellate dalle altre in un ordine di somministrazione all'interno di ogni epoca; ciò spiega il ricorso ad una *token frequency* pari a 5 anche nell'addestramento bilanciato. In ogni caso lo stato generale della mappa è influenzato dalle parole mostrate in precedenza attraverso l'apprendimento a lungo termine (*storage*). Durante l'apprendimento, le regole hebbiane vengono applicate a entrambi i livelli di connettività, in modo che i nodi che sono altamente reattivi a un determinato stimolo diventino sempre più reattivi a quello stesso stimolo, mentre i nodi che sono debolmente reattivi a uno stimolo, lo diventino ancora meno. In particolare, i pesi su tutte le connessioni di ingresso a $BMU(t)$ sono regolati per essere più vicini al segnale di ingresso corrente; allo stesso modo, tutte le connessioni temporali a $BMU(t)$ sono regolate per essere maggiormente correlate con il pattern di attivazione complessivo della mappa al tempo $t-1$; vale a dire, la connessione da $BMU(t-1)$ a $BMU(t)$ è potenziata, mentre le connessioni da tutti i nodi tranne $BMU(t-1)$ a $BMU(t)$ sono depresse. La regolazione del peso si diffonde radialmente ai nodi vicini con una funzione gaussiana centrata sull'attuale BMU. La propagazione radiale richiede la condivisione delle informazioni e la dipendenza dall'addestramento tra nodi topologicamente adiacenti, che sono quindi addestrati a rispondere allo stesso modo a stimoli di input simili (Pirrelli, Ferro e Marzi 2015). La rete elabora e organizza i dati, in base all'emergere di caratteristiche che possano raggrupparli, con l'obiettivo di modellare una dinamica adattiva evinta dalle proprietà strutturali dell'input. Dunque, l'organizzazione generale di una TSOM sarà determinata dalla struttura morfologica dei dati di training, in base a tre fattori: similarità tra forme di input, frequenza delle stesse, e contesto dei simboli. Ogni forma di parola in ingresso è rappresentata da una serie temporale di simboli somministrati alla TSOM uno alla volta. L'input sequenziale mantiene aggiornato lo stato di attivazione in

funzione degli «eventi» passati per ogni simbolo di input, mentre la dinamica ricorrente (Elman 2004) prevede ogni nodo connesso con il layer di input e tutti gli altri nodi della mappa (incluso se stesso) attraverso connessioni temporali rientranti con ritardo temporale di 1. Più specificamente, alla presentazione di un simbolo sul layer di input, tutti i nodi della mappa vengono attivati contemporaneamente attraverso le loro connessioni spaziali e temporali, in una dinamica a breve e a lungo termine. Sul breve termine, al tempo t uno stimolo di input viene mostrato alla mappa codificato sul livello di input. Le informazioni si propagano attraverso le connessioni ai nodi della mappa, attivati in modo sincrono in funzione della forza dei pesi della connessione. Nel frattempo, gli stessi nodi ricevono segnali rientranti attraverso le connessioni, aggiornando lo stato di attivazione del nodo all'istante corrente (BMU_t) con il livello di attivazione dei nodi che rispondono allo stimolo immediatamente precedente (tempo $t-1$). La competizione tra i nodi co-attivati è vinta dal nodo più altamente attivato, definito come *Best Matching Unit* (BMU) al tempo t e i pesi delle connessioni temporali — in aggiornamento dello stato di attivazione — codificano quanto il nodo maggiormente attivo all'istante corrente (BMU_t) è atteso dalla BMU_{t-1} . In termini di correlazione temporale, i nodi vincenti che si attivano serialmente si connettono sempre più fortemente durante l'apprendimento, in funzione della frequenza di input. La forza delle connessioni va a definire il livello di radicamento di una certa forma nel lessico e può essere interpretata dinamicamente come la probabilità condizionata con cui si prevede che una particolare forma si manifesti, quando viene percepita una parte sempre più lunga della parola. Il potenziamento del peso si diffonde ai nodi più prossimi all'attuale BMU, in modo che quelli topologicamente vicini diventino sempre più sensibili a stimoli simili sui due di connettività, favorendo una specializzazione mentre sono inibiti i nodi distanti dalla BMU. Stando a una dinamica di organizzazione spaziale, i nodi vicini tendono ad attivarsi e a competere per input simili. A favorire la specializzazione e il radicamento, tra l'altro, sono le parole ad alta frequenza e isolate che tendono ad attivare catene di nodi dedicate, mentre parole a bassa frequenza in famiglie «numerose» tendono a co-attivare catene condivise (o parzialmente condivise). Questa prospettiva rende possibile misurare le relazioni associative tra input correlati e il contributo di queste relazioni all'apprendimento. Per la mappa, le parole sono sequenze di simboli che iniziano con "#" e terminano con "\$" e che vengono mostrati uno dopo l'altro sul livello di input nell'ordine da sinistra a destra. Un

simbolo della parola corrente attiva una BMU e il processo viene ripetuto per ogni simbolo fino a raggiungere quello di fine parola ("\$"). Lo stato di attivazione della mappa viene eventualmente azzerato alla presentazione di una nuova parola, riconosciuta con "#" sul livello di input. Osservando la struttura morfologica emergente è evidente una correlazione diretta tra la segmentazione morfologica e l'organizzazione topologica delle BMU sulla mappa: le parole di input che condividono costituenti sub-lessicali tendono a attivare catene di nodi prossimi tra loro o fortemente co-attivate tra loro, per cui si può definire la vicinanza topologica sulla mappa e il livello di co-attivazione come una funzione della prossimità morfologica.

3.3.3 Accuratezza: *recoding e recall*

Alla fine di ogni epoca di addestramento la «risposta» della mappa è nell'attivazione dei nodi.

L'accuratezza nel ricodificare correttamente ogni simbolo delle forme di input – *recoding* – verifica se per ogni forma di input le BMU sono associate correttamente con i simboli di input.

Il compito di verificare se una forma di input è memorizzata correttamente può essere misurato come la capacità di «richiamare» (*recall*) correttamente la traccia memorizzata delle forme di input. Il *recall* è modellato come il compito di ripristinare una forma subito dopo l'esposizione alla stessa, quindi un protocollo sperimentale che evidenzia l'interazione dinamica tra integrazione/supporto a breve termine delle tracce mnestiche e la memorizzazione a lungo termine di informazioni lessicali. Questo tipo di misurazione è fortemente legato agli effetti di frequenza per cui il confronto tra la percezione della struttura (e quindi la capacità di anticipare il simbolo all'istante $t+1$ dato il simbolo all'istante t) con l'andamento dell'accuratezza nel recupero della traccia seriale dall'input sincronico, può offrire notevoli evidenze nel contrastare una frequenza bilanciata (i.e., simulando una ideale distribuzione uniforme delle frequenze) con una distribuzione di frequenze realistiche o *corpus-based* nell'interazione con la struttura morfologica emergente. Esaminando le catene di attivazione delle BMU è possibile stabilire se la mappa riconosce o meno l'input in ingresso come una specifica sequenza di simboli. Poiché, si ricorda, ogni forma è esposta alla mappa una lettera alla volta, ogni simbolo lascia uno schema di attivazione nel buffer a breve termine della mappa. Quando viene mostrato l'intero modulo, viene mantenuta l'attivazione simultanea di tutte le lettere che compongono la parola appena vista. In definitiva, quando una serie temporale di simboli di input viene terminata, il

modello Integrated Activation Pattern (IAP, ovvero una rappresentazione sincronica dei nodi che si sono serialmente attivati per quella parola di input)⁹ risultante dei nodi rappresenta la risposta di elaborazione della mappa all'intera serie di input. Questa informazione può essere usata come pattern di attivazione dell'input per testare quanto bene la mappa può effettuare una recall della parola dal suo pattern. Una parola viene richiamata correttamente se tutti i suoi simboli vengono richiamati correttamente nell'ordine appropriato. Il recall di una parola dalla sua IAP può essere più o meno facile a seconda del grado di co-attivazione di altre IAP non target le cui BMU sono altamente attivate nella IAP target. Ad esempio, se due stringhe di input presentano alcuni simboli in comune, tenderanno ad attivare schemi di nodi ampiamente sovrapposti.

3.3.4 Predizione

La capacità della rete di anticipare una forma di input può essere valutato incrementalmente come la predizione per il simbolo corrente dato i simboli precedenti. Per ogni BMU correttamente pre-attivata si assegna il valore di 1 incrementato per i valori dei simboli precedenti. Si assegna il valore 0 per un simbolo non anticipato correttamente. Più simboli sono anticipati e più è facile la predizione di una parola, o di una sotto-sequenza di essa.

Una efficace strategia di predizione della sequenza di una parola suggerisce una maggiore facilità di elaborazione della stessa.

⁹La somma vettoriale dei modelli di attivazione attivati da ciascun simbolo nella parola.

Capitolo 4

ANALISI DEI RISULTATI

Questo capitolo è dedicato alla presentazione degli esperimenti condotti addestrando la rete TSOM¹, in diverse condizioni di somministrazione e costituzione dell'input, e ad un'attenta analisi quantitativa delle dinamiche di apprendimento della rete e della sua organizzazione finale, tramite il software *R* (<https://cran-r.project.org>).

Il comportamento della TSOM è stato osservato prima nella condizione di esposizione a 600 forme derivate, 200 per ogni suffisso derivazionale *-zione*, *-tore*, *-ico*, in una distribuzione bilanciata (ogni singola forma è stata somministrata alla TSOM con una frequenza pari a 5). Questa distribuzione innaturale costituisce la base dell'ESPERIMENTO 1.

Nell'ESPERIMENTO 2 le stesse forme derivate sono somministrate alla TSOM con le frequenze derivate da corpus.

Infine, l'ESPERIMENTO 3 intende valutare il ruolo della famiglia di forme derivate e flessi riconducibili alla stessa base, per un sottoinsieme delle forme derivate selezionate per gli esperimenti 1 e 2.

¹Gli esperimenti sono stati condotti nell'ambito del *ComphysLab* dell'istituto di Linguistica Computazionale, CNR (<http://www.comphyslab.it/>).

4.1 Le dinamiche di *learning* e *processing*

Nel primo esperimento è stata addestrata una rete TSOM di 45×45 nodi con 600 forme derivate², con una distribuzione uniforme di frequenze. L'ordine con il quale le forme derivate sono somministrate alla rete è casuale per ognuna delle 100 epoche di addestramento. In modo del tutto arbitrario è stato scelto di somministrare ogni singola forma con una frequenza di 5 *token*. Ogni sessione di addestramento è stata ripetuta per 5 volte con i medesimi parametri, così da mediare i risultati delle ripetizioni e controllare eventuali comportamenti casuali delle singole sessioni.

Per simulare effetti per lo più strutturali delle dinamiche di *learning* e *processing* e il loro impatto su un'organizzazione morfologica coerente, durante il training sono state fornite solo informazioni relative alla forma e alla frequenza. Ogni parola di input è stata somministrata alla TSOM a un simbolo alla volta, per cui la memoria dei simboli già somministrati è di fatto ripristinata alla presentazione del simbolo di inizio di input "#", seguendo una rappresentazione ortogonale³, come ad esempio riportato qui di seguito:

$$\# = [1 0 0 0 0 \dots]$$

$$A = [0 1 0 0 0 \dots]$$

$$B = [0 0 1 0 0 \dots]$$

$$I = [0 0 0 1 0 \dots]$$

$$T = [0 0 0 0 1 \dots]$$

Le parole somministrate come stimoli seriali, ripetuti con una certa frequenza per ogni epoca di addestramento, portano la rete a consolidare il suo stato di attivazione modificando il peso delle connessioni tra nodi e trasformando quella che può essere una risposta occasionale in uno schema di attivazione sistematicamente associato a quello stimolo lessicale. Per questa dinamica di addestramento, l'apprendimento di una parola consiste nella trasformazione adattiva di una risposta occasionale all'input.

Specificamente, l'elaborazione dell'input lessicale coinvolge più sotto-processi: l'attivazione di nodi specifici in "risposta" alla somministrazione di stimoli (nel nostro caso sequenze di simboli ortografici) ricodificati internamente dalla rete (*recoding* o ricodifica dei simboli di

²Si veda l'elenco delle forme riportato in Appendice.

³In Appendice è riportata la codifica completa.

input); il recupero, dalla memoria a lungo termine della mappa, della serie temporale corrispondente alla rappresentazione di ogni parola di input (*recall* o recupero della traccia di memoria); lo sviluppo delle aspettative data una serie temporale di simboli di input (*predizione*). Pertanto a fine addestramento è stata valutata l'accuratezza sia per il *recoding* che *recall*. L'accuratezza del compito di *recoding* quantifica l'abilità della TSOM di ricodificare correttamente ogni simbolo delle forme di input⁴. Una forma è ricodificata correttamente se tutte le BMU sono associate correttamente con i simboli di input. Il compito di *recall*, invece, permette di testare quanto una mappa è in grado di “recuperare” la sequenza di attivazione dei nodi della mappa a partire da un pattern di attivazione sincronico, ovvero quanto accuratamente l'attivazione dei nodi della mappa ricodifica l'informazione temporale delle catene di simboli somministrati in input.

La predizione consente, infine, di monitorare l'abilità della mappa di predire una forma di input, mediante la valutazione incrementale del numero dei simboli della forma di input che la mappa riesce correttamente ad anticipare. Per ogni BMU correttamente pre-attivata si calcola incrementalmente il valore di 1, viceversa 0. Più simboli di input vengono anticipati e più facile è la predizione per quella parola. Una efficace strategia di predizione della sequenza di una parola target non solo suggerisce una maggiore facilità di *processing* seriale di quella forma, ma quantifica, in un certo senso, anche l'aspettativa globale per certe sequenze tipiche dato un insieme di forme di input.

Successivamente all'addestramento relativo all'Esperimento 1, sono quantificati *recoding* e *recall*. Si vedano i risultati mediati sulle 5 ripetizioni riportati in Tabella 4.1.

<i>compito</i>	<i>accuratezza in %</i>	<i>deviazione standard in %</i>
recoding	100	0
recall	97.6	0.27

Tabella 4.1: Esperimento 1: percentuali di accuratezza per *recoding* e *recall*.

⁴Il compito di *recoding*, apparentemente semplice nel caso presente – con una accuratezza pari al 100% per ogni esperimento – porta la rete a sviluppare una specializzazione per i simboli che costituiscono l'input. Nel caso, ad esempio, di esposizione a lessici di lingue diverse, per epoche diverse di addestramento, porterebbe la rete a sviluppare una corretta ricodifica solo per i simboli della lingua ai quali è stata sufficientemente esposta.

Gli errori di ricodifica si hanno quando un simbolo in ingresso attiva una BMU associata ad un simbolo diverso, mentre errori di *recall* possono verificarsi quando la mappa richiama in modo errato uno o più simboli nella stringa di input, sostituendolo con un simbolo diverso o richiamando simboli nell'ordine sbagliato, oppure quando avviene un *recall* parziale della parola target (come ad esempio #,A,F,I,L,I,A,Z,I,O,N,E,\$ per l'input #,A,F,F,I,L,I,A,Z,I,O,N,E,\$).

In Figura 4.1 sono riportati gli andamenti per il compito di *recoding* e *recall* durante le epoche di addestramento. Nel dettaglio, la ricodifica è rappresentata dalla curva blu il cui andamento definisce quando e quanto correttamente viene ricodificato un simbolo di input. Sull'asse delle y è riportata la percentuale di accuratezza con la variabilità sulle 5 ripetizioni indicata dai "baffi" verticali. Come si può osservare, il compito di ricodifica è più semplice di quello di *recall*, e lo precede nelle epoche di addestramento in quanto è – in un certo senso – funzionale al *recall*. Infatti, affinché una forma possa essere recuperata dalla memoria a lungo termine della rete è necessario che la ricodifica interna alla rete per quella sequenza di simboli ricodifichi correttamente i simboli di input.

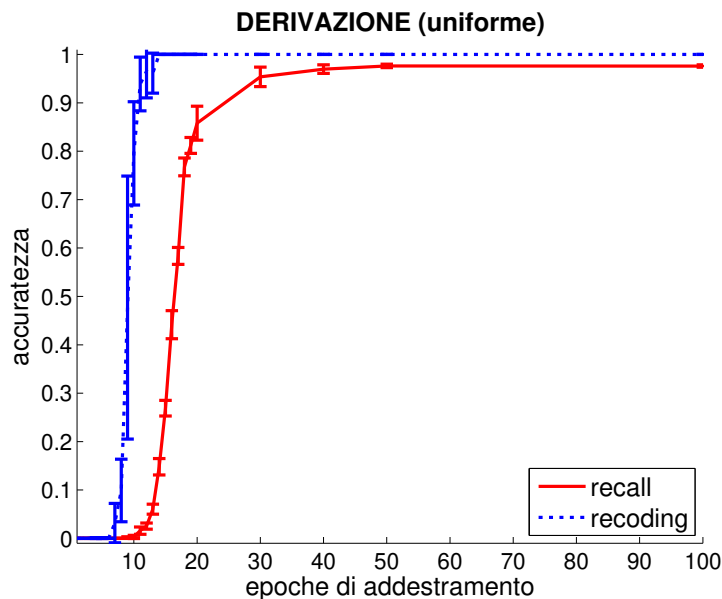


Figura 4.1: Recoding (curva blu) e Recall (curva rossa) nell'addestramento con frequenze bilanciate (Esperimento 1).

Il *recall*, rappresentato dalla curva rossa, raggiunge infatti un buon livello di accuratezza solo a partire dall'epoca 20 e dalla 30 si stabilizza sui valori di accuratezza vicina a quelli della

fine dell'addestramento.

Poiché non tutte le forme derivate riportano il suffisso derivazionale più frequente ma un suo allomorfo, per ognuno dei tre suffissi *-zione*, *-tore*, *-ico* considero separatamente proprio i derivati che presentano un'allomorfia. In Figura 4.2 sono riportate le curve di accuratezza per il solo compito di *recall* per derivati e derivati-allomorfi.

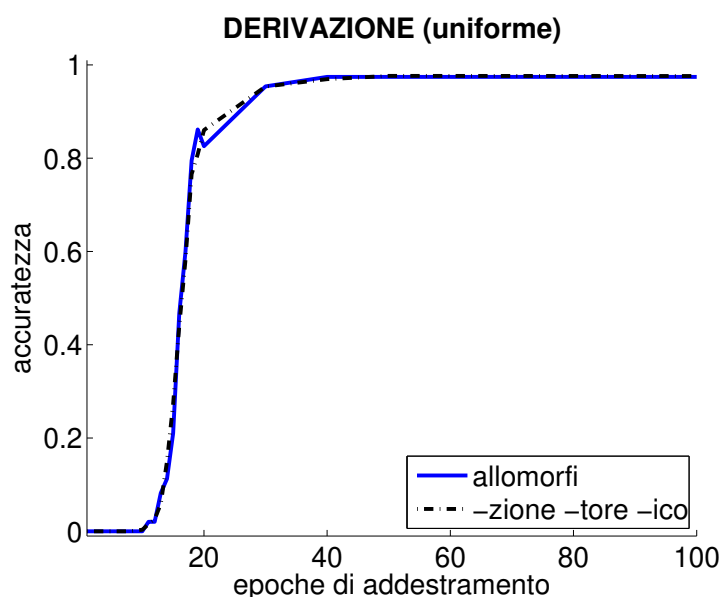


Figura 4.2: Recall per le forme derivate con suffissi *-zione*, *-tore*, *-ico* (curva tratteggiata nera) e per le forme derivate con allomorfi (curva continua blu) nell'addestramento con frequenze bilanciate (Esperimento 1).

Come si può notare, tra l'epoca 20 e l'epoca 32, l'accuratezza nel recupero delle tracce di memoria corrispondenti alle catene temporali dei derivati che presentano una allomorfia nel suffisso è leggermente inferiore, anche se statisticamente significativa.

Uno degli obiettivi degli esperimenti è monitorare il modo in cui la TSOM percepisce la struttura morfologica delle forme derivate. Il compito di *predizione* consente di monitorare le dinamiche di accesso lessicale, analizzando gli effetti della predicibilità delle forme derivate sulla (auto-)organizzazione morfologica delle rete.

La Figura 4.3 riporta per ognuna delle serie derivazionali (i.e. forme derivate in *-zione*, *-tore*, *-ico*) l'andamento della predizione all'epoca finale (100): ogni forma derivata è stata allineata con le altre sul primo simbolo del suffisso (si veda sull'asse $x=0$ il punto di confine di morfema). Come misura della temporalità e della serialità dell'input ho codificato nel *dataframe* la distanza dal confine di morfema, pertanto i suffissi derivazionali per ogni forma sono rappresentati nei plot dai valori di $x \geq 0$. Viceversa, le basi delle forme corrispondono ai valori negativi sull'asse delle x .

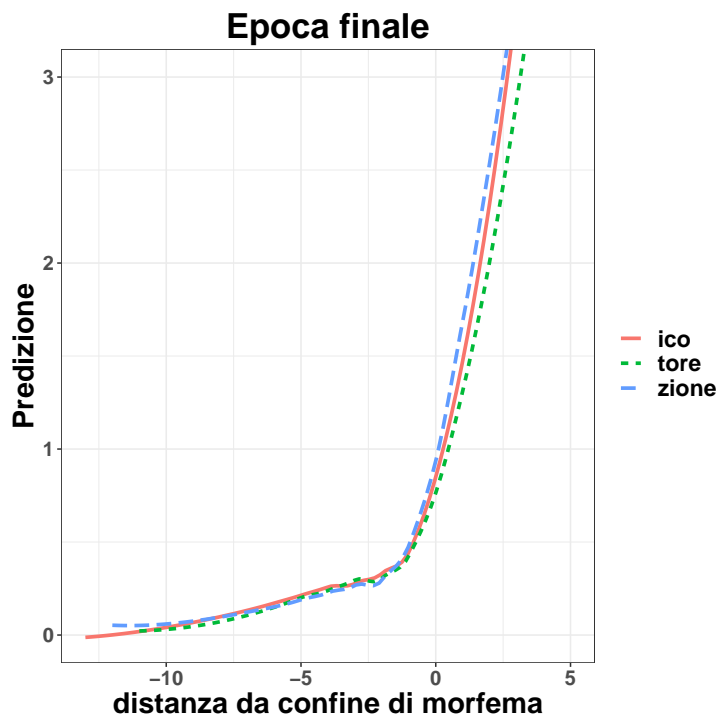


Figura 4.3: Predizione per le forme derivate in *-zione*, *-tore*, *-ico* all'epoca finale (100).

Come si può ben osservare, la rete riesce a sviluppare una ottima capacità predittiva per le sequenze di simboli relative ai suffissi (si ricorda che per ogni suffisso derivazionale sono somministrate alla rete 200 forme). Per la parte radicale, invece, la rete non riesce a sviluppare molte aspettative probabilistiche, sia perché ogni forma è presentata con la medesima frequenza, sia perché la rete è addestrata con 600 forme diverse tra loro. A completamento del plot riporto un modello di regressione non lineare (funzione *gam* di R)⁵ per l'interazione tra la posizione nella sequenza seriale dei simboli e le serie dei suffissi come variabile categoriale, con le ripetizioni dell'addestramento come variabile casuale. Il modello è applicato all'epoca finale (si veda la Tabella 4.2 riferita al plot in Figura 4.3).

Parametric coefficients	<i>Edf</i>	<i>Std.Error</i>	<i>t-value</i>	<i>p-value</i>
Intercept <i>-zione</i>	0.610576	0.015987	38.192	0.00201
Intercept <i>-tore</i>	0.051188	0.016568	3.090	< 2e-16
Intercept <i>-ico</i>	1.312900	0.017634	74.452	< 2e-16
dist. to MB <i>-zione</i>	0.172507	0.004285	40.257	< 2e-16
dist. to MB <i>-tore</i>	0.039346	0.004548	8.651	< 2e-16
dist. to MB <i>-ico</i>	0.261068	0.003352	77.875	< 2e-16
$R^2(\text{conditional}) = 57.1\%$				

Tabella 4.2: ESPERIMENTO 1. Coefficienti e significatività relative al modello non lineare *gam* per l'epoca finale (*predizione del simbolo* \sim *distanza da confine di morfema* * *suffisso* + *s(ripetizione.id, bs = "re"), epoca finale*)).

Si evidenzia una differenza significativa nella capacità predittiva (si noti in Tabella 4.2 la colonna relativa al *p-value*) per *-zione*, *-tore* e *-ico* che seppur di moderata entità (si vedano i coefficienti per le intercette e le pendenze) indica una predizione sistematicamente maggiore

⁵<https://cran.r-project.org/> Poiché i fenomeni naturali hanno in genere andamenti curvilinei, posseggono quindi punti di massimo o minimo, flessi e asintoti, più difficilmente possono essere descritti con funzioni lineari. Per questo motivo la funzione *gam* viene applicata a set di dati che hanno una relazione non lineare tra variabili dipendenti e indipendenti.

per i derivati in *-zione* rispetto a quelli in *-tore*, e ancor più in *-ico*. Indicativo della robustezza del modello *gam* è un R^2 particolarmente alto (0.57) che spiega il 57% della varianza dei dati.

Dopo aver evidenziato la dinamica della predizione all'epoca finale, è interessante spostare l'attenzione alle epoche iniziali di addestramento. Infatti, come evidenziato in Figura 4.1, le epoche critiche nelle quali la mappa stabilizza la sua organizzazione topologica e nelle quali determina le connessioni temporali sono le epoche iniziali (10-30). Pertanto, in Figura 4.4 riporto la curva di regressione per la predizione delle forme suddivise nei 3 suffissi per le epoche 10, 20, 30, 40, 50 confrontate con l'epoca finale (100).

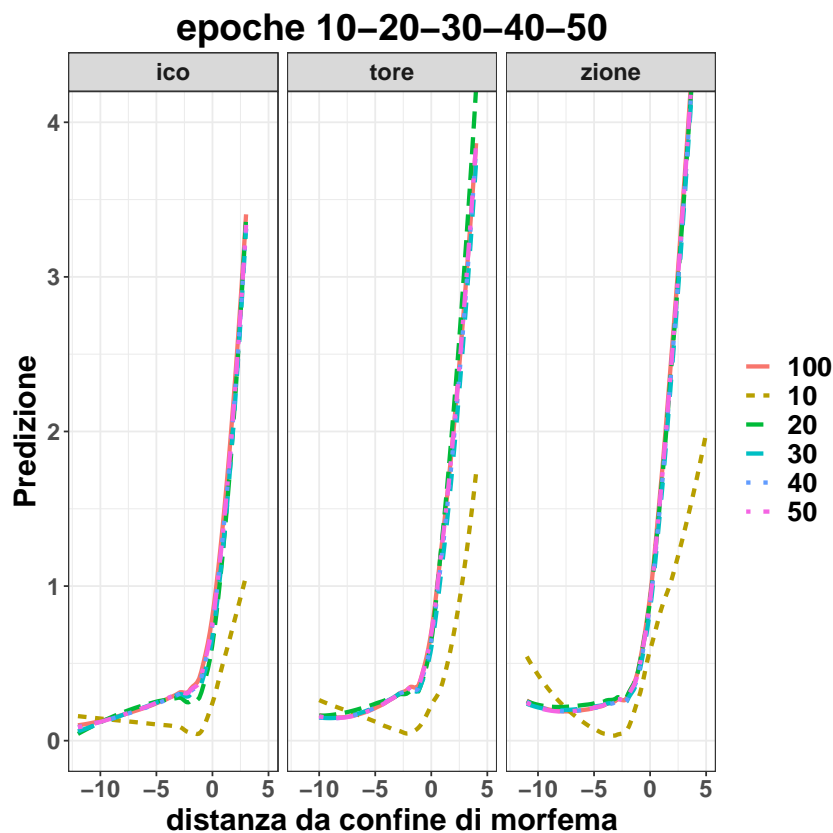


Figura 4.4: Predizione per le epoche 10, 20, 30, 40, 50 confrontate con l'epoca finale (100) per i derivati suddivisi per suffisso.

Nonostante la rete all'epoca 20 non sia in grado di recuperare la traccia di memoria di tutte le forme – o meglio, di tutte quelle che riesce a richiamare correttamente alla fine dell'addestramento – è ugualmente in grado di percepire la struttura interna delle forme di input.

La curva di regressione per l'epoca 20 è coincidente con quella delle epoche successive, compresa l'epoca finale. Selezionando un range ancora più specifico delle epoche intermedie, è possibile evidenziare in modo più dettagliato la dinamica evolutiva della percezione della struttura morfologica delle forme. Di fatto, differenze sostanziali per i tre suffissi sono visibili - in riferimento alla Figura 4.5 - tra le epoche 10 e 12, con una precocità individuabile per il suffisso più saliente; all'epoca 14 gli andamenti sono simili.

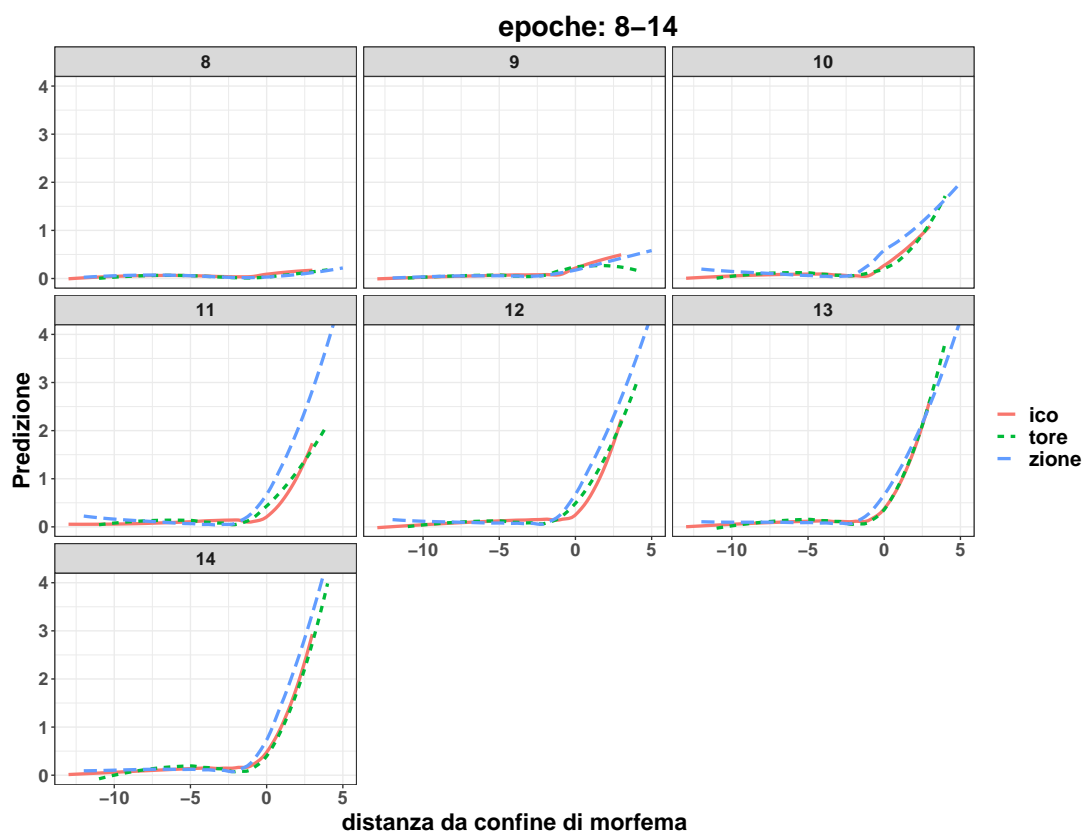


Figura 4.5: Predizione per le epoche 8, 9, 10, 11, 12, 13, 14 per le tre serie di suffissi.

Questa ultima analisi suggerisce che la percezione della struttura morfologica di forme complesse e ridondanti, come quelle che costituiscono l'input della rete, è dissociata dalla loro memorizzazione e capacità di recupero della traccia di memoria. L'evidenza sembra suggerire che la percezione della struttura interna di forme morfologicamente complesse sia, in un certo senso, preliminare e funzionale alla loro memorizzazione più efficace, proprio per il ruolo che le sotto-strutture ridondanti possono svolgere per agevolare il compito di memorizzazione.

4.2 Il ruolo della frequenza dell'input

Volendo evidenziare un diverso profilo di *processing* dell'input in funzione del grado di predizione dello stimolo e in base alla ridondanza delle strutture ricorrenti, è fondamentale indagare nel dettaglio gli effetti della frequenza dell'input valutando il ruolo effettivo della *token frequency*.

Pertanto, per l'Esperimento 2 la rete è stata addestrata sullo stesso set di forme somministrate nell'Esperimento 1 in funzione della frequenza con la quale le forme occorrono nel Corpus *DerIvaTario*: alla TSOM sono state sottoposte le 600 forme nell'intervallo di frequenza 2-999.

Anche in questo caso ogni sessione di allenamento è stata ripetuta 5 volte e le percentuali di accuratezza sono state mediate per 5 ripetizioni per controllare la variabilità nell'ambito delle singole sessioni. Si ricorda che nell'addestramento della TSOM, i pesi di connessione sono modulati dalla distribuzione degli input secondo i principi hebbiani, quindi tendenzialmente ci si aspetterebbe che un input ad alta frequenza attivando ripetutamente lo stesso pattern di nodi rafforzi le connessioni tra i nodi ad attivazione sequenziale, rendendo le parole ad alta frequenza associate a pattern di attivazione altamente reattivi. Le parole ad alta frequenza sviluppano un rapido radicamento delle connessioni *hebbiane* che alla fine favoriscono alti livelli di attivazione dei nodi nelle loro tracce di memoria e una co-attivazione più sparsa delle tracce di memoria di altre parole. Connessioni forti e livelli di attivazione elevati significano aspettative elevate per le tracce di memoria attivate di frequente, che vengono quindi richiamate più facilmente e sono meno confondibili con altre parole vicine.

Bisogna tuttavia tenere conto del fatto che le parole ad alta frequenza non necessariamente condividano con altre forme le sotto-strutture morfologiche che le compongono, ed ancora che gli effetti di alta *token frequency* possano portare ad effetti di "offuscamento" della percezione delle stesse; tutto ciò potrebbe rallentare le dinamiche di *recall*. L'accuratezza per la ricodifica dei simboli di input e di recupero della serie temporale è evidenziata (sempre in termini percentuali di accuratezza per *recoding* e *recall*) nei risultati mediati sulle 5 ripetizioni, riportati in Tabella 4.3.

<i>compito</i>	<i>accuratezza in %</i>	<i>deviazione standard in %</i>
recoding	100	0
recall	96,06	1,429

Tabella 4.3: Esperimento 2: percentuali di accuratezza per *recoding* e *recall*.

Anche nel caso di somministrazione con una distribuzione non bilanciata – *corpus-based* –, il compito di ricodifica anticipa quello di *recall* nelle epoche di addestramento, ma questo ultimo raggiunge un buon livello di accuratezza solo a partire dall’epoca 30, stabilizzandosi dalla 50 fino a fine addestramento (figura 4.6).

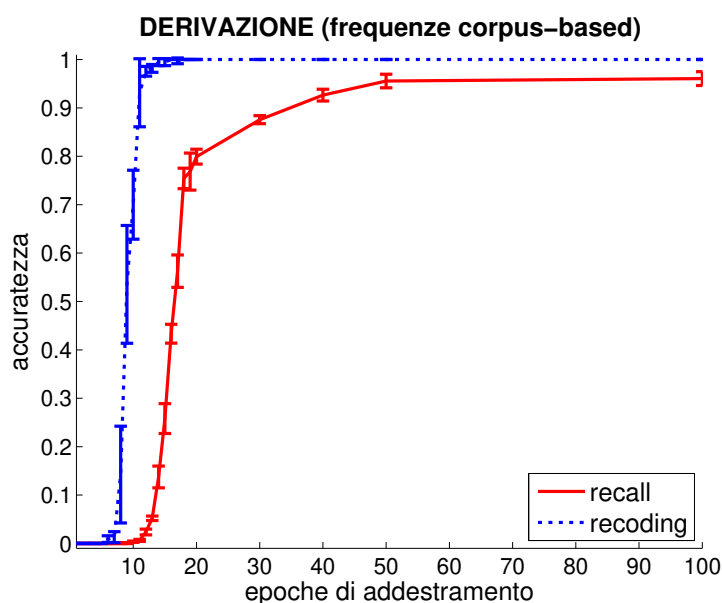


Figura 4.6: Recoding (curva blu) e Recall (curva rossa) nell’addestramento con frequenze da corpus (Esperimento 2).

In aggiunta, sempre considerando in maniera separata le forme derivate che presentano un’allomorfia, riporto le curve di accuratezza per il solo compito di recall in Figura 4.7): dopo l’epoca 20 il recupero delle tracce di memoria per le catene temporali con derivati che presentano una allomorfia nel suffisso è visibilmente inferiore, complici le frequenze nettamente più basse degli allomorfi.

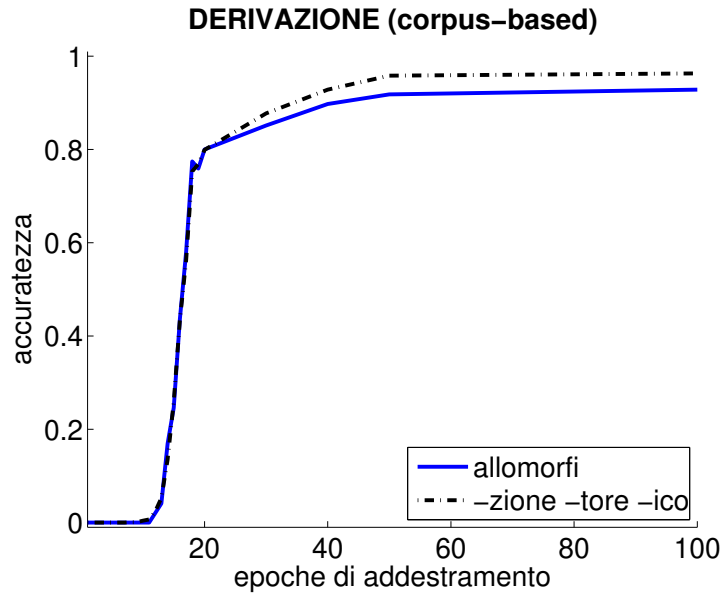


Figura 4.7: Recall per le forme derivate con suffissi *-zione*, *-tore*, *-ico* (curva tratteggiata nera) e per le forme derivate con allomorfi (curva continua blu) nell'addestramento con frequenze da corpus (Esperimento 2)

Per verificare il modo in cui la TSOM percepisce la struttura morfologica delle forme derivate in base a frequenze da corpus, il compito di *predizione* è esemplificativo. La Figura 4.8 riporta per le serie derivazionali in esame l'andamento della predizione all'epoca finale (100). Le diverse distribuzioni di frequenza portano un effetto di maggiore emergenza delle basi (per valori negativi sull'asse x) rispetto all'addestramento con frequenze bilanciate (Esperimento 1). La frequenza, in questo senso, sembra essere particolarmente importante nel determinare in maniera considerevole l'abilità della rete nel compito predittivo, poiché la rete stessa dimostra di dipendere fortemente dalla ridondanza dell'input, che favorisce la memorizzazione e l'anticipazione delle sequenze più frequenti.

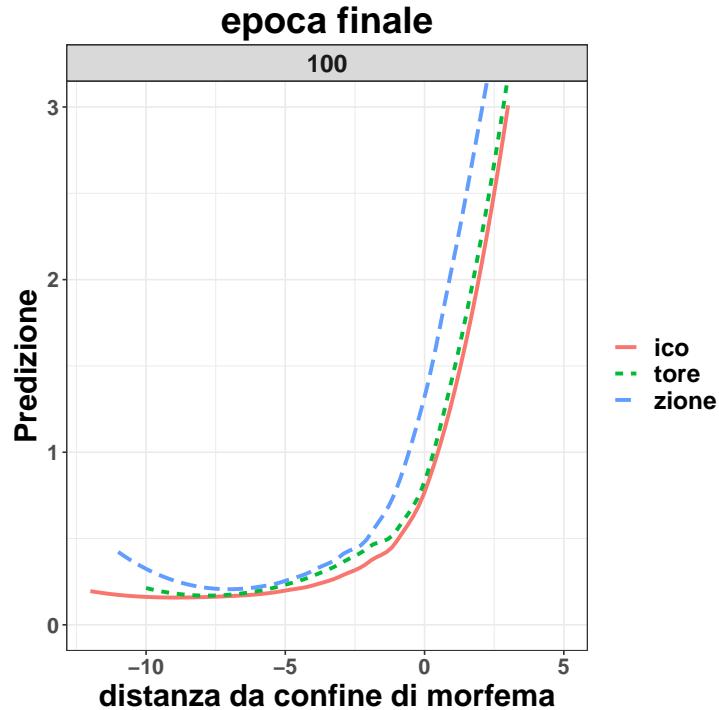


Figura 4.8: Predizione per le forme derivate in *-zione*, *-tore*, *-ico* all'epoca finale (100)

Confrontando i modelli additivi a effetti misti (*gam*), senza interazione della frequenza e con interazione della frequenza, riportati rispettivamente nelle Tabelle 4.4 e 4.5, in base alla distanza da confine morfema (*dist.to.MB*)⁶, è possibile notare come l'intercetta arrivi ad assumere misure relativamente alte per i valori $x > 0$; ciò indica una capacità di predizione più alta: si vedano i valori riportati in Tabella 4.5.

Il modello riportato in Tabella 4.5, a conferma degli andamenti esplicitati dai plot di regressione, suggerisce quindi un'incidenza rilevante della frequenza sulla capacità predittiva della rete, in quanto sembra agevolarla in modo significativamente non dissimile per le tre serie in esame (si veda il p-value). La varianza complessivamente spiegata dal modello è pari al 55.3% delle osservazioni, come indicato dal coefficiente R^2 .

⁶La *slope* è indicativa di quanto cresce la predizione per la forma derivata, fino alla sue fine.

	<i>Edf</i>	<i>Std.Error</i>	<i>t-value</i>	<i>p-value</i>
Intercept <i>-zione</i>	0.979535	0.018153	53.96	< 2e-16
Intercept <i>-tore</i>	0.373210	0.018813	19.84	< 2e-16
Intercept <i>-ico</i>	1.286010	0.014032	91.65	< 2e-16
dist. to MB <i>-zione</i>	0.223025	0.004866	45.84	< 2e-16
dist. to MB <i>-tore</i>	0.101199	0.005164	19.60	< 2e-16
dist. to MB <i>-ico</i>	0.231551	0.003807	60.83	< 2e-16
$R^2(\text{conditional}) = 53.8\%$				

Tabella 4.4: Coefficienti e significatività relative al modello non lineare *gam* applicato all'epoca finale, senza interazione della frequenza (*predizione del simbolo ~ distanza da confine di morfema * suffisso + s(ripetizione.id, bs = "re"), epoca finale*)

Parametric coefficients	<i>Edf</i>	<i>Std.Error</i>	<i>t-value</i>	<i>p-value</i>
Intercept <i>-zione</i>	8.369e-01	1.835e-02	45.60	< 2e-16
Intercept <i>-tore</i>	3.452e-01	1.852e-02	18.64	< 2e-16
Intercept <i>-ico</i>	1.209e+00	1.399e-02	86.44	< 2e-16
dist. to MB <i>-zione</i>	2.224e-01	4.786e-03	46.47	< 2e-16
dist. to MB <i>-tore</i>	9.989e-02	5.080e-03	19.66	< 2e-16
dist. to MB <i>-ico</i>	2.315e-01	3.744e-03	61.83	< 2e-16
frequency	2.559e-03	7.613e-05	33.62	< 2e-16
$R^2(\text{conditional}) = 55.3\%$				

Tabella 4.5: Coefficienti e significatività relative al modello non lineare *gam* applicato all'epoca finale (*predizione del simbolo ~ distanza da confine di morfema * suffisso + frequenza + s(ripetizione.id, bs = "re"), epoca finale*)

..

Appurata l'incidenza della *token frequency*, è utile mettere a fattore se le minime – seppur significative - differenze tra i 3 suffissi siano riconducibili o meno anche alla diversa lunghezza del suffisso (5, 4, 3, rispettivamente per *-zione, -tore, -ico*), per cui è stata inclusa come variabile nel modello riportato in Tabella 4.9

	<i>Edf</i>	<i>Std.Error</i>	<i>t-value</i>	<i>p-value</i>
Intercept <i>-zione</i>	0.503435	0.135201	3.724	0.000197
Intercept <i>-tore</i>	0.042796	0.152881	0.280	0.779532
Intercept <i>-ico</i>	0.608358	0.087989	6.914	4.79e-12
dist. to MB <i>-zione</i>	0.222611	0.004861	45.792	< 2e-16
dist. to MB <i>-tore</i>	0.100518	0.005161	19.476	< 2e-16
dist. to MB <i>-ico</i>	0.231551	0.003803	60.888	< 2e-16
serie*lunghezza (<i>-zione</i>)	0.010293	0.028710	0.359	0.719958
serie*lunghezza (<i>-tore</i>)	0.029211	0.035856	0.815	0.415269
serie*lunghezza (<i>-ico</i>)	0.225884	0.029626	7.625	< 2.51e-14
$R^2(\text{conditional}) = 53.8\%$				

Tabella 4.6: Coefficienti e significatività relative al modello non lineare *gam* applicato all'epoca finale (*predizione del simbolo ~ distanza da confine di morfema * suffisso + serie * lunghezza del suffisso + s(ripetizione.id, bs = "re"), epoca finale*)

Ciò che risulta evidente dalle misure riportate è che tali differenze non sembrano dipendere dalla diversa lunghezza dei tre suffissi; anche la precocità dapprima attribuita alla salienza di *-zione* è fondamentalmente relativa alla sua sequenza ortotattica, in quanto unità di elaborazione morfologica meno confondibile con sequenze non corrispondenti al suffisso. In effetti, il trigramma “zio” è poco o per niente attestato in altri contesti di parola (limitatamente all'input somministrato alla rete, ma non molto dissimile dalla completezza del lessico a cui può essere esposto un parlante italiano) e quindi fortemente connotativo, cosa che rende subito individuabile il confine e quindi l'intera struttura compositiva⁷.

⁷Nella word-prediction, generalmente, l'analisi distribuzionale dei contesti linguistici rivela le preferenze

A dare effettiva riprova di questa intuizione sono i risultati di un ulteriore modello, riportato in Tabella 4.7, in cui ho contrastato le allomorfie limitatamente alla serie *-zione* per fornire un quadro di dettaglio nel contesto della serie esaminata, considerata più saliente.

	<i>Edf</i>	<i>Std.Error</i>	<i>t-value</i>	<i>p-value</i>
Intercept <i>-zione</i>	1.76169	0.13309	13.237	< 2e-16
Intercept <i>-gione</i>	0.74962	0.13505	5.551	2.90e-08
Intercept <i>-ione</i>	1.45748	0.13808	10.555	< 2e-16
dist.to.MB <i>-zione</i>	0.30089	0.03984	7.553	4.55e-14
dist.to.MB <i>-gione</i>	0.16066	0.03967	4.050	5.15e-05
dist.to.MB <i>-ione</i>	0.25177	0.04129	6.098	1.11e-09
$R^2(\text{conditional}) = 59.1\%$				

Tabella 4.7: Coefficienti e significatività relative al modello non lineare *gam* applicato all'epoca finale (*predizione del simbolo* \sim *distanza da confine di morfema * allomorfo + s(ripetizione.id, bs = "re"), epoca finale*)).

Ciò che viene messo in luce in maniera ancora più dettagliata è il ruolo pressoché ininfluente della lunghezza del suffisso, rispetto alla sequenza dello stesso, per cui *-zione* - come sequenza ortotattica - o fonotattica - raffrontato non solo agli altri suffissi presi in esame, ma anche ai suoi allomorfi - induce una capacità predittiva significativamente maggiore nella rete.

Come per le analisi effettuate per il primo esperimento, dopo aver mostrato la dinamica nel compito di predizione all'epoca finale (figura 4.9), andando a ritroso fino alle epoche iniziali di addestramento è possibile trarre evidenze considerevoli su quali siano quelle cruciali - in relazione ai due livelli di connessione - per l'individuazione ed eventualmente la condivisione della struttura. L'incremento per le catene di attivazione relative alle rappresentazioni dei suffissi suggerisce che la predizione riflette l'organizzazione paradigmatica delle parole nella mappa. A questo proposito, se la possibilità di riscontrare una leggera anticipazione sullo combinatorie delle parole. A un certo livello teorico, alcune potenzialità combinatorie sono determinate da tratti morfo-sintattici e semantici generali delle parole stesse, in quanto membri di una classe linguistica astratta.

stem condiviso - per effetti di frequenza - si può palesare all'epoca finale, non è invece possibile riscontrarla precocemente in fase di emergenza del suffisso per le esigue epoche iniziali (per questo motivo si rivela interessante indagare ulteriormente il ruolo della famiglia paradigmatica, verificando la percezione della struttura per un sottoinsieme di forme derivate riconducibili alla stessa base).

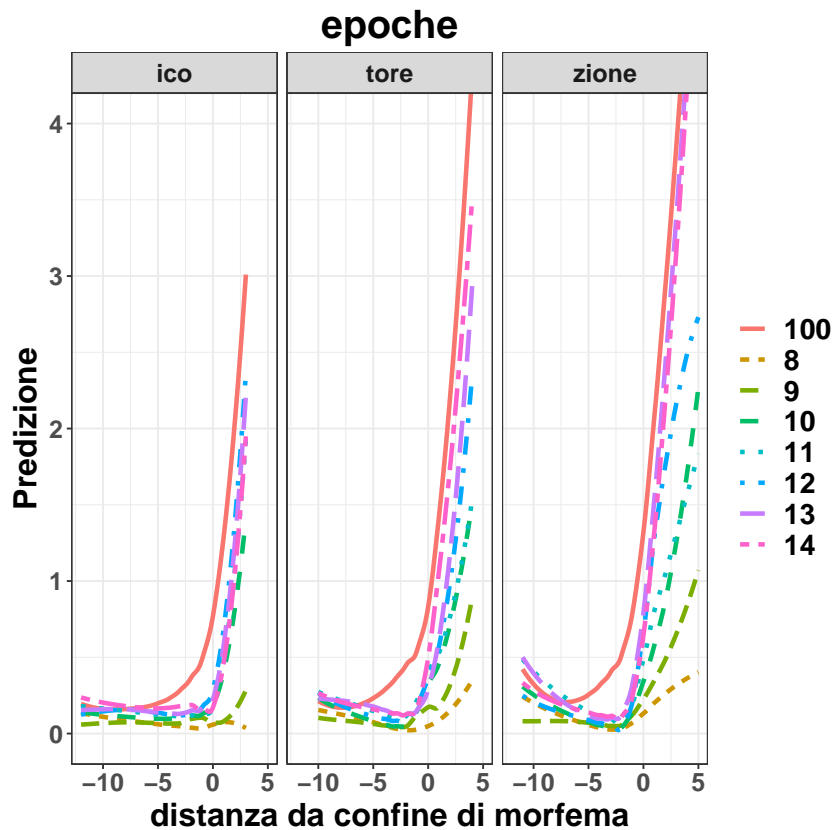


Figura 4.9: Predizione delle epoche intermedie confrontate con l'epoca finale (100) per i derivati suddivisi per suffisso.

In virtù delle osservazioni classificate per suffisso, la Figura 4.10 riproduce in maniera più chiara le differenze negli andamenti delle tre serie per le epoche dalla 8 alla 14.

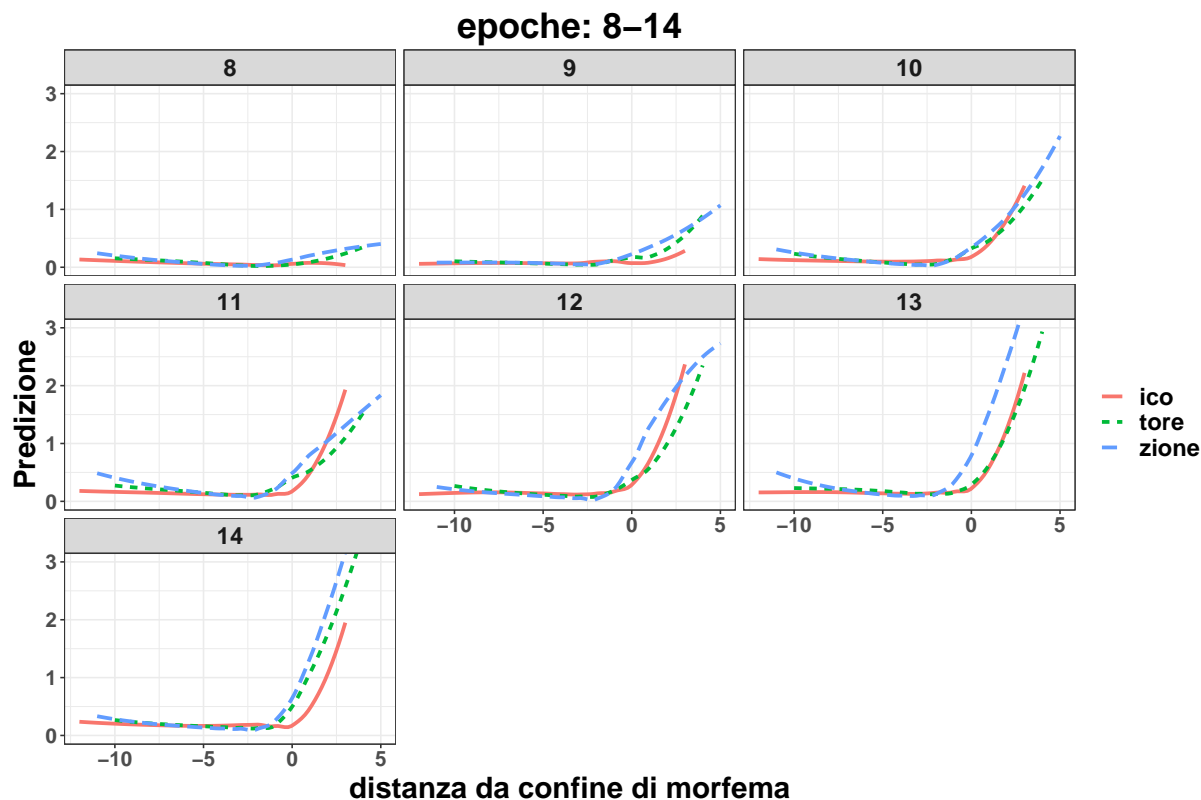


Figura 4.10: Predizione per le epoche 8, 9, 10, 11, 12, 13, 14.

I risultati evidenziati, a questo punto, decretano un'importante connessione tra la struttura morfologica, le distribuzioni di frequenza delle parole e la memorizzazione competitiva di forme. L'acquisizione di pattern morfologici ridondanti gioca un ruolo sempre più importante in un lessico emergente, ponendo il focus sulle strategie di acquisizione basate sull'integrazione dinamica di forme.

Sulla base di questi assunti e sulle difficoltà elaborative per la parte radicale delle forme derivate, ha senso chiedersi se possano emergere con maggiore facilità per le forme con frequenze maggiori. Per questo motivo sono considerati cinque gruppi per frequenze rispettivamente basse, medio-basse, medie, medio-alte, alte:

- Nel primo gruppo rientrano le frequenze da 2 a 10,
- nel secondo gruppo quelle dal primo quartile alla mediana,
- il terzo gruppo si identifica nel terzo quartile,

- il quarto gruppo considera le frequenze oltre il terzo quartile ed entro i 100,
- il quinto gruppo comprende tutte le frequenze al di sopra di 100.

In un confronto diretto tra le epoche precoci e quella finale, nei plot che seguono (Figure 4.11 e 4.12), il progressivo emergere della struttura interna delle forme derivate evidenzia da un lato la precocità della parte suffissale in epoche di addestramento precoci – che quindi ancora non vedono la memorizzazione delle forme piene – mentre dall’altro un’incapacità elaborativa per la parte tematica. Specificamente alle epoche che vanno dalla 8 alla 14 non sono riscontrati effetti della frequenza nell’emergenza delle basi, come invece accade all’epoca finale (100) alla fine dell’addestramento (come evidenziato in Figura 4.12).

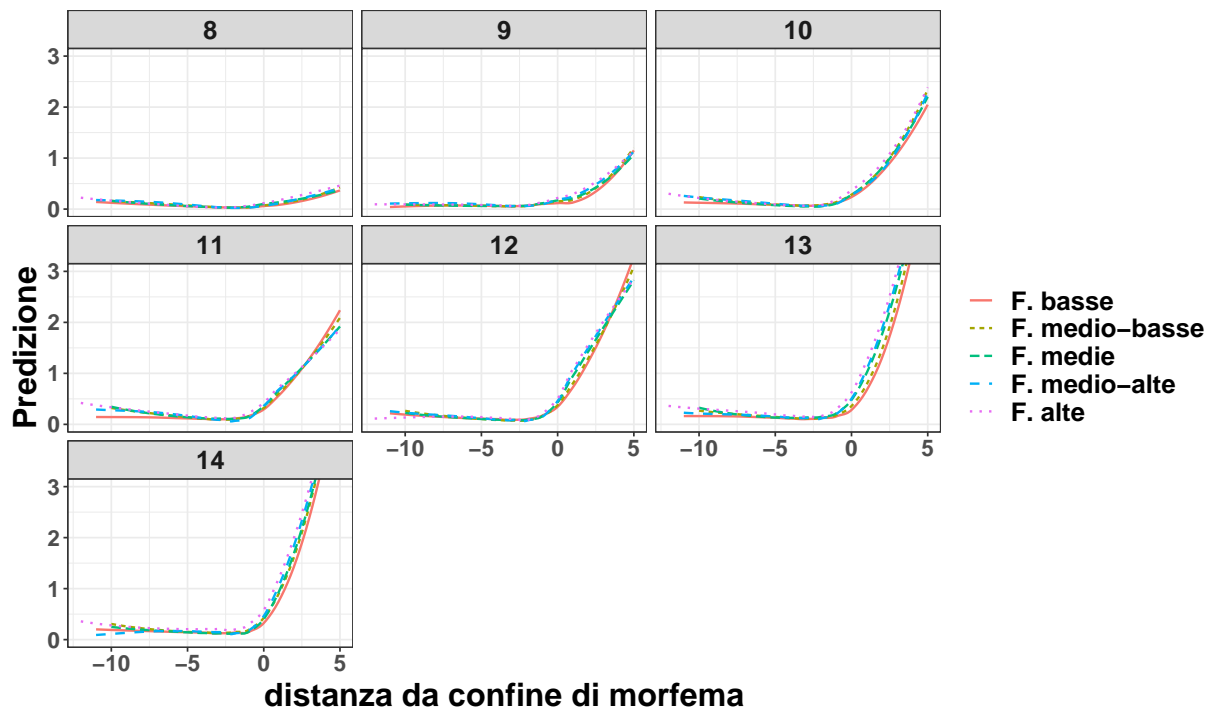


Figura 4.11: Predizione per gruppi di frequenze per le serie in *-ico*, *-tore*, *-zione* alle epoche 8, 9, 10, 11, 12, 13, 14.

All'epoca finale (figura 4.12), per gruppi di frequenze crescenti, si osserva invece un progressivo effetto di radicamento anche delle basi. Inoltre, per le forme rientranti nel gruppo delle frequenze alte, la predizione della rete cresce in modo quasi lineare mano a mano che ci si avvicina alla fine di parola, favorendo la specializzazione e il radicamento di catene di nodi dedicate.

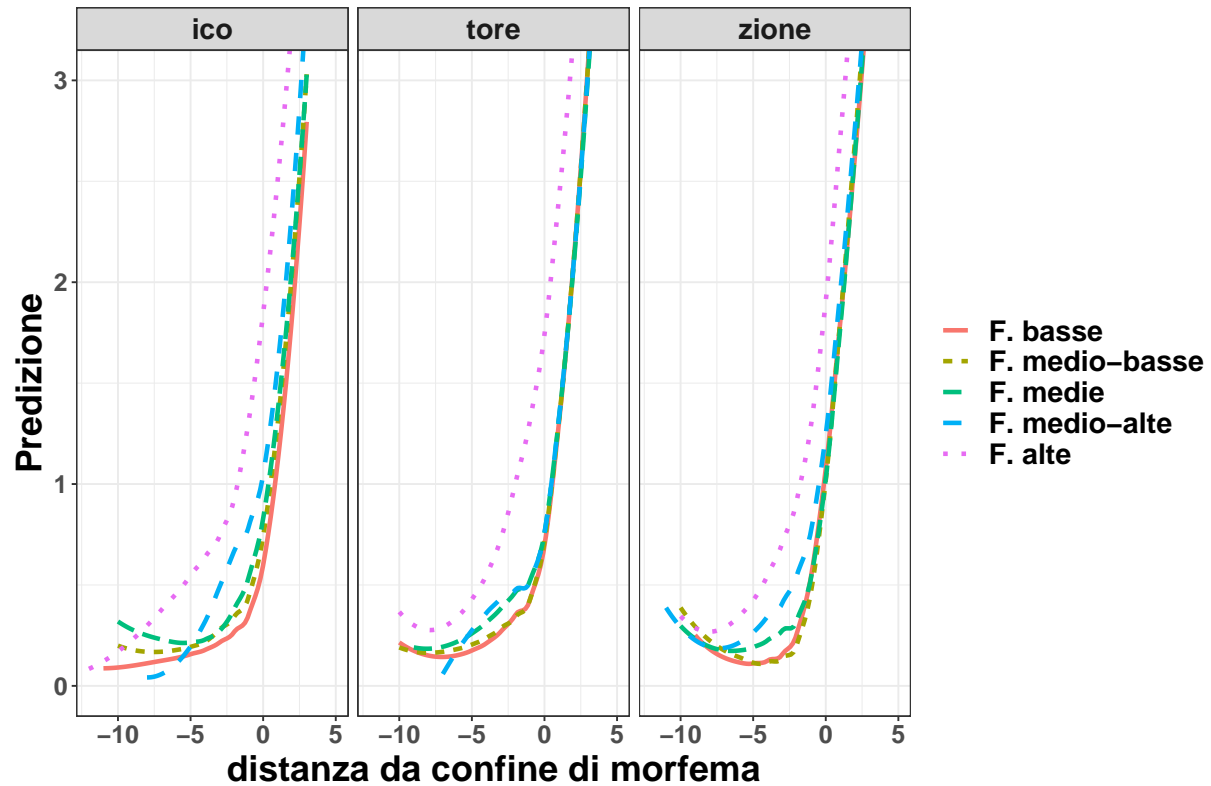


Figura 4.12: Predizione per gruppi di frequenze per le serie in *-ico*, *-tore*, *-zione* all'epoca finale.

4.3 Il ruolo della famiglia

Con l'obiettivo di controllare gli effetti dell'interazione tra *token frequency* e *type frequency* sulla dinamica di memorizzazione ed elaborazione della rete ho pensato ad un esperimento con una selezione dei dati che, per un numero limitato delle forme derivate, vede l'aggiunta sia di forme derivate con altri suffissi che forme flesse appartenenti alla stessa famiglia – ovvero riconducibili alla stessa base⁸.

Nello specifico, ho selezionato 94 forme con i suffissi derivazionali *-zione*, *-tore*, *-ico* e 506 altre forme morfologicamente complesse – forme derivate e forme flesse della base verbale stessa – per un insieme 57 famiglie.

Per questo Esperimento 3, ho utilizzato per l'addestramento sempre una rete TSOM di 45x45 nodi di memoria. I valori di accuratezza mediati sulle 5 ripetizioni per *recoding* e *recall* sono riportati in Tabella 4.8.

<i>compito</i>	<i>accuratezza in %</i>	<i>deviazione standard in %</i>
recoding	100	0
recall	85,92	0,915

Tabella 4.8: Esperimento 3: percentuali di accuratezza per *recoding* e *recall*.

In Figura 4.13 si riporta, ancora una volta, l'andamento per *recoding* e *recall* in chiave evolutiva. Come si può osservare, per questo insieme di dati di input la rete non riesce ad incrementare ulteriormente – tra le epoche 20 e 50 circa – l'accuratezza nella sua capacità di recupero delle tracce seriali dato il pattern sincronico di ogni parola. Questo è dovuto ad una dinamica di memorizzazione COMPETITIVA tra forme suffissate in *-zione*, *-tore*, *-ico* e i molti altri suffissi possibili (*-ale*, *-bile*, *-evole*, *-mente*, *-oso*, *-trice* per le forme derivate, e *-are*, *-iamo*, *-ando*, *-endo* per citare solo alcuni dei suffissi verbali attestati).

A ciò si aggiunge anche una potenziale competizione tra le allomorfe della base. Se da una parte la presenza nell'input di più forme riconducibili alla stessa base aumenta l'esposizione alla base stessa – nel caso di famiglie regolari – dall'altra parte la presenza di allomorfi –

⁸Si veda l'elenco delle forme ordinate per famiglie in Appendice.

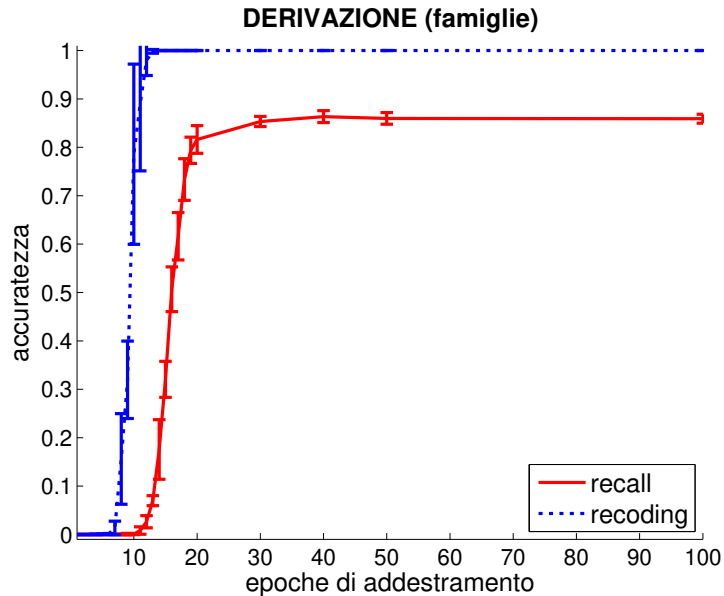


Figura 4.13: Recoding (curva blu) e Recall (curva rossa) per tutte le forme derivate e flesse nell'addestramento con famiglie (Esperimento 3).

nel caso di famiglie irregolari – fa aumentare la competizione tra le sequenze di nodi che si attivano.

Questo, in una dinamica seriale di forme memorizzate contestualmente, porta ad aumentare l'incertezza della rete per il simbolo da attivare (all'istante $t+1$) dato il simbolo all'istante corrente (BMU_t , si veda §3.3).

In Figura 4.14 si riporta le epoche iniziali (8:14) la predizione per le forme derivate in *-zione*, *-tore*, *-ico* contrapposte a tutte le altre forme derivate e flesse. Come si può osservare, gli "altri" suffissi – diversi da quelli oggetto di studio della presente tesi – difficilmente riescono ad essere anticipati dalla rete, per l'alto numero di suffissi-tipo che ogni base può attivare.

Il quadro non cambia radicalmente se osserviamo lo stesso quadro all'epoca finale (si veda la Figura 4.15).

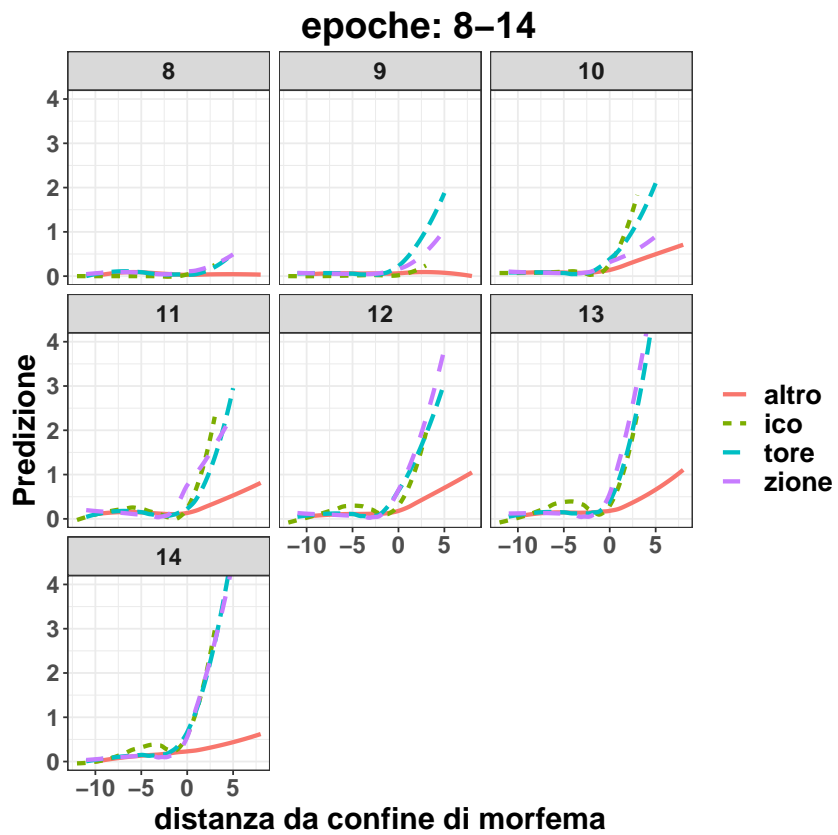


Figura 4.14: Predizione per le forme derivate in *-zione*, *-tore*, *-ico* contrapposte a tutte le altre forme ("altro") per le epoche iniziali (8:14).

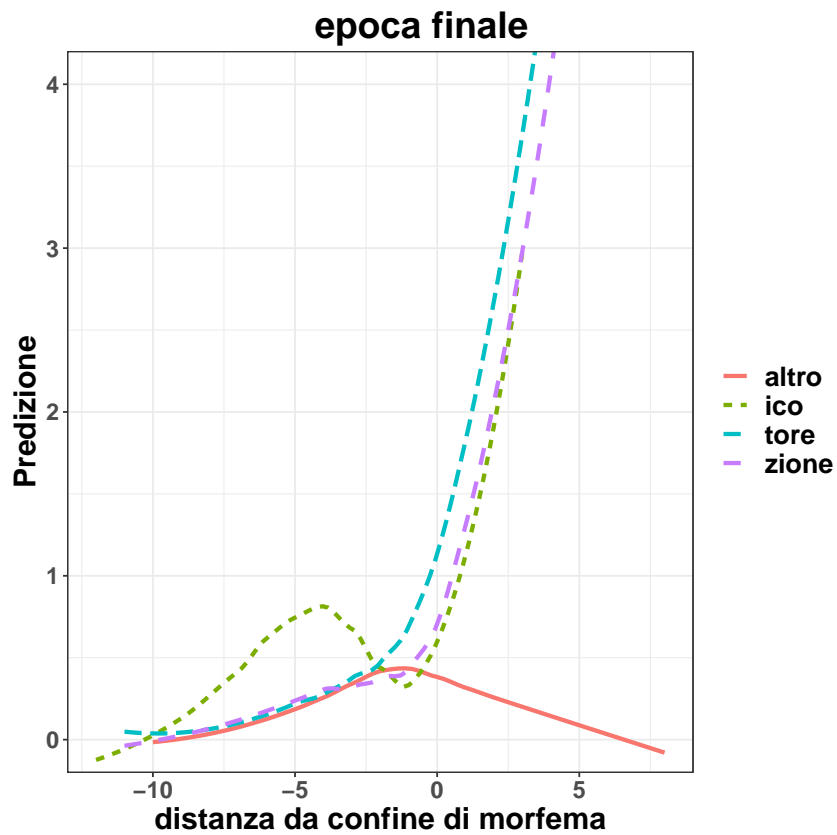


Figura 4.15: Predizione per le forme derivate in *-zione*, *-tore*, *-ico* contrapposte a tutte le altre forme ("altro") all'epoca finale (100).

Il modello non lineare relativo al plot di Figura 4.15 – con un R^2 che spiega il 46.1% della varianza dei dati – significativamente evidenzia la predizione maggiore per le forme dei tre suffissi rispetto a tutte le forme suffissate – derivate e flesse – etichettate come "altro".

	<i>Edf</i>	<i>Std.Error</i>	<i>t-value</i>	<i>p-value</i>
Intercept <i>-zione</i>	1.536812	0.014655	104.868	< 2e-16
Intercept <i>-tore</i>	1.694607	0.018035	93.962	< 2e-16
Intercept <i>-ico</i>	0.986764	0.069896	14.118	< 2e-16
Intercept (" <i>altro</i> ")	0.286146	0.007326	39.057	< 2e-16
dist. to MB <i>-zione</i>	0.331543	0.003519	94.215	< 2e-16
dist. to MB <i>-tore</i>	0.337043	0.004340	77.653	< 2e-16
dist. to MB <i>-ico</i>	0.133367	0.013442	9.922	< 2e-16
dist. to MB " <i>altro</i> "	0.011894	0.001340	8.875	< 2e-16
$R^2(\text{conditional}) = 46.1\%$				

Tabella 4.9: Coefficienti e significatività relative al modello non lineare *gam* applicato all'epoca finale (*predizione del simbolo ~ distanza da confine di morfema * suffisso + serie * lunghezza del suffisso + s(ripetizione.id, bs = "re"), epoca finale*)).

La maggiore predizione per le basi – in questo esperimento – è ricondotta agli effetti della *type frequency*. Poiché la derivazione italiana avviene principalmente attraverso l'agglutinazione, più grande è la famiglia e maggiormente sarà facilitata l'attivazione di nodi per le sequenze più frequenti (o produttive). A delucidare questo fenomeno ed esemplificare il plot riportato in Figura 4.15, sono le poche forme suffissate in *-ico*: scenico (scena), informatico (informare), illusionistico (illudere).

Ancora una volta si è rivelato opportuno tenere in considerazione la lunghezza come fattore di input. Poiché il dataset di riferimento per l'esperimento 3 evidenzia una lunghezza media delle forme maggiore rispetto al set di dati utilizzato per gli esperimenti precedenti, è plausibile assumere che possa avere un peso considerevole nell'influenzare dinamicamente l'emergere delle strutture morfologiche. Questo aspetto è stato indagato riproponendo – nel modello non lineare *gam* – la lunghezza come predittore (Tabella 4.10).

	<i>Edf</i>	<i>Std.Error</i>	<i>t-value</i>	<i>p-value</i>
Intercept <i>-zione</i>	1.4830132	0.0153293	96.744	< 2e-16
Intercept <i>-tore</i>	1.6601790	0.0181069	91.688	< 2e-16
Intercept <i>-ico</i>	0.9579703	0.0695160	13.781	< 2e-16
Intercept " <i>altro</i> "	1.45748	0.13808	10.555	< 2e-16
dist.to.MB <i>-zione</i>	0.3502183	0.0036887	94.943	< 2e-16
dist.to.MB <i>-tore</i>	0.3488879	0.0043633	79.960	< 2e-16
dist.to.MB <i>-ico</i>	0.1614263	0.0134438	12.007	< 2e-16
dist.to.MB " <i>altro</i> "	0.0975749	0.0060179	16.214	< 2e-16
dist.to.MB*lunghezza	-0.0086098	0.0005962	-14.441	< 2e-16
$R^2(\text{conditional}) = 46.9\%$				

Tabella 4.10: Coefficienti e significatività relative al modello non lineare *gam* applicato all'epoca finale con effetto della lunghezza delle forme (*predizione del simbolo* \sim *distanza da confine di morfema * suffisso + distanza da confine di morfema * lunghezza + s(ripetizione.id, bs = "re"), epoca finale*)).

Come si può osservare dai coefficienti riportati per i predittori considerati nel modello, la lunghezza porta ad una riduzione della capacità predittiva della rete, nonostante la misura stessa di predizione sia agevolata dalla stessa – in virtù della sua natura incrementale.

Poiché non tutte le forme derivate e suffissate contengono in modo trasparente la base, è opportuno considerare questo effetto di (ir)regolarità della famiglia nelle mie analisi.

Riporto a tal fine l'andamento delle curve di accuratezza per il compito di *recall* per le epoche di addestramento separatamente per le forme in famiglie regolari contrapposte a quelle in famiglie irregolari (si veda Figura 4.16).

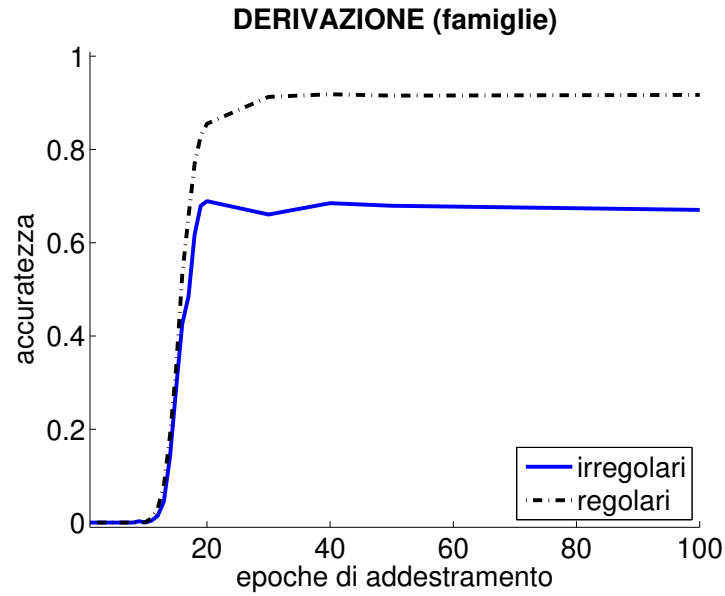


Figura 4.16: Recall per le forme derivate regolari (curva tratteggiata nera) e irregolari (curva continua blu) nell’addestramento con famiglie (Esperimento 3).

A partire dall’epoca 20 il compito di *recall* per le forme irregolari resta stabilizzato su livelli significativamente più bassi. In linea di massima le forme regolari vengono richiamate più facilmente rispetto alle irregolari, dove c’è una maggiore competizione tra le sequenze di nodi che si attivano. Più precisamente, in una condizione iniziale uniforme, ogni forma flessa o derivata attiva la stessa sequenza di nodi per la base che è trasparentemente percepibile nelle famiglie regolari. Altresì, per le famiglie irregolari, le forme flesse e derivate attivano sia la base che suoi allomorfi, riducendo sia la memorizzazione che la predizione della stessa, sia per una minore predicibilità dell’allomorfo sia per la condivisione/attivazione della stessa base o allomorfo da parte di un numero minore di *word-type*.

Come si può osservare in Figura 4.17 la capacità predittiva della rete per le forme che "regolarmente" attivano tutte la stessa base è maggiore rispetto a quanto non riesce a fare per le forme che presentano, all’interno della famiglia, la compresenza di più allomorfi della base. In Tabella 4.11 è riportato il modello *gam* con l’inclusione della lunghezza come predittore, e delle famiglie come effetto casuale (*random effect*).

La presenza di una sola base "regolare" aumenta il numero di forme che attivano la stessa sequenza di nodi – *type frequency effect*. A ciò si contrappone l’attivazione COMPETITIVA di più

(allomorfi) basi da parte di forme della stessa famiglia: come, ad esempio, nel caso di *aggredire*, *aggressione*, *aggressore*, *illudere*, *illusione*, *illusionista*, *leggere*, *letto*, *lettore*, ecc.

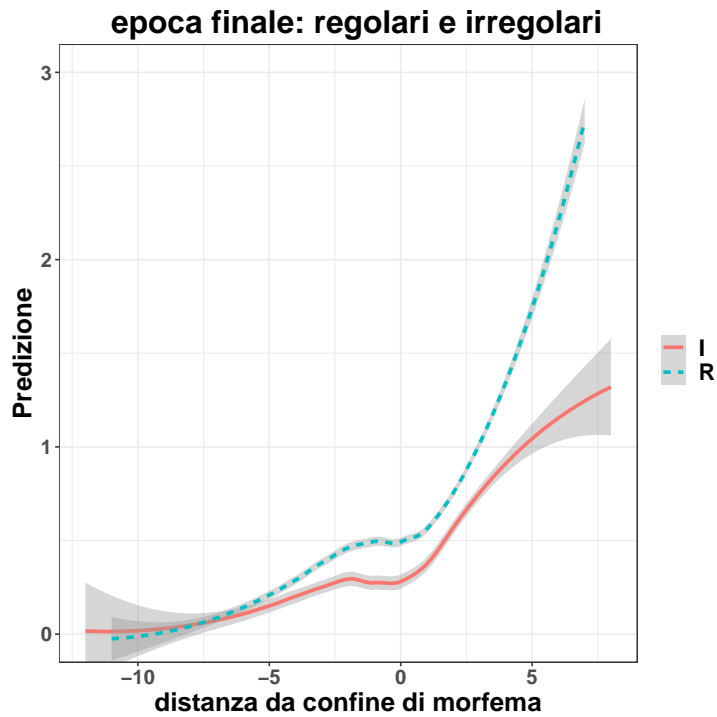


Figura 4.17: Predizione per le forme derivate in famiglie regolari (linea tratteggiata verde) e irregolari (linea continua rossa) (Esperimento 3). Le aree grigie indicano l'intervallo di confidenza del modello.

	<i>Edf</i>	<i>Std.Error</i>	<i>t-value</i>	<i>p-value</i>
Intercept Irregolari	-0.141685	0.084731	-1.672	0.09450
Intercept Regolari	-0.404719	0.096524	-4.193	2.76e-05
Lunghezza Irregolari	0.071089	0.007669	9.269	< 2e-16
Lunghezza Regolari	0.051355	0.008702	5.902	3.64e-09
dist.to.MB Irregolari (<i>slope</i>)	0.028669	0.014333	2.000	0.04548
dist.to.MB Regolari (<i>slope</i>)	-0.083141	0.016808	-4.947	7.59e-07
dist.to.MB*lunghezza Irregolari	0.004564	0.001404	3.251	0.00115
dist.to.MB*lunghezza Regolari	0.009861	0.001629	6.053	1.43e-09
$R^2(\text{conditional}) = 15.6 \%$				

Tabella 4.11: Coefficienti e significatività relative al modello non lineare *gam* applicato all'epoca finale con effetto della lunghezza delle forme regolari e irregolari (*predizione del simbolo* \sim *distanza da confine di morfema * IR * lunghezza + s(famiglia, bs = "re"), epoca finale*)).

Capitolo 5

DISCUSSIONE E CONSIDERAZIONI FINALI

5.1 Percezione della struttura morfologica

L'analisi dei processi di memorizzazione e percezione della struttura, alla base dell'obiettivo posto in essere nella presente tesi, si è focalizzato sul fenomeno dei derivati in italiano, nello specifico delle parole in *-zione*, *-tore*, *-ico*. In particolare, sono partita dall'ipotesi di una possibile diversa salienza tra i suffissi in esame, come caso di studio per evidenziare strategie elaborative che prendono avvio dal fenomeno della percezione della struttura morfologica dei parlanti come base della competenza lessicale.

Attraverso l'osservazione – e modellazione – del comportamento di una rete neurale artificiale ricorrente (come lo è la *Temporal Self-Organising Map*) della predizione, del tutto dinamica in funzione dei fattori distribuzionali, è stato possibile individuare un correlato della percezione della struttura morfologica dell'input del parlante, basata sulla ridondante presenza nell'input di strutture ricorrenti in più contesti. Il tipo di rete proposto per le simulazioni si è rivelato estremamente efficace nel delineare la dinamica dell'emergenza della struttura morfologica in funzione delle caratteristiche distribuzionali opportunamente individuate per i tre esperimenti: (i) una ideale distribuzione bilanciata di sole forme derivate (in uno stesso numero di tipi per ognuna delle serie derivazionali), (ii) una distribuzione *corpus-based* delle stesse forme, infine (iii) una distribuzione bilanciata di forme derivate sia nelle tre serie in

esame che altre, unitamente a forme flesse della base. Questa ultima condizione sperimentale ha consentito di evidenziare l'effetto della *type-frequency*, avendo tenuto fermo l'effetto della *token-frequency*.

Le successive analisi quantitative del comportamento della rete, seppur con una base comune delle dinamiche di memorizzazione e di predizione, mostrano una certa variabilità delle strategie elaborative proprio in funzione delle diverse condizioni di esposizione all'input.

In chiave evolutiva, il progressivo emergere della struttura interna delle forme derivate, con un precoce emergere proprio della parte suffissale in epoche di addestramento che ancora vedono la rete non aver memorizzato le forme piene, offre evidenza sperimentale nel quadro della riflessione teorica discussa nella parte iniziale di questa tesi.

L'incertezza elaborativa per la parte tematica delle forme suffissate (si vedano i livelli prossimi allo 0 per la predizione nei *regression plots* del Capitolo 4), contrapposti ai livelli che epoca dopo epoca tendono a incrementare per le catene di attivazione relative alle rappresentazioni dei suffissi, suggerisce che la predizione riflette l'organizzazione paradigmatica delle parole nella mappa. Infatti, quanto maggiore è la cardinalità della *serie* derivazionale, tanto più alto è il livello della predizione in corrispondenza del suffisso, e dunque più facile l'elaborazione dello stesso.

Suggerisce, ancora, l'indipendenza della percezione di sotto-strutture lessicali (coincidenti, in questo e nei casi di paradigmi flessionali, con strutture morfemiche) dalla loro posizione nella forma piena stessa. Come il bambino che nelle prime fasi di acquisizione riesce a memorizzare prima la parte suffissale di alcuni verbi che seguono lo schema flessionale più frequente (e.g. *ato* per "mangiato"), anche la rete dimostra di dipendere fortemente dalla ridondanza dell'esposizione all'input, memorizzando e riuscendo ad anticipare ciò che è più frequente.

A ciò segue un progressivo processo di memorizzazione che riflette una dinamica di apprendimento strutturale da parte della rete. La capacità di una TSOM di predire la struttura dell'input corrente non è altro che il risultato finale di un processo ADATTIVO di auto-organizzazione delle catene temporali sviluppate come rappresentazione interna alla rete in risposta alle parole di input.

La diversa consistenza funzionale dei tre suffissi *-zione*, *-tore*, *-ico* (Alessandro Laudanna e Burani 1995) emerge nelle minime differenze che si osservano nei modelli statistici, confer-

mando che la rete è in grado di percepire le differenze distribuzionali e sviluppare una coerente capacità predittiva. Per un parlante, una forma che termina in *-zione* è quasi esclusivamente una forma derivata, così come il 78% delle forme terminanti in *-tore* e il 52% in *-ico* (Giraudo e Dal Maso 2016, in base alle distribuzioni del corpus *CoLFIS*). Allo stesso modo, per la rete la probabilità di occorrenza degli *ngrammi* corrispondenti ai suffissi modula la loro salienza percettiva. Nel dettaglio, l'aspettativa – e con essa la capacità predittiva – della rete per la sequenza *-zione* dipende strettamente dalla probabilità di *incontrare* una "z" nell'input, e la stessa per i simboli successivi "ione" dipenderà dalla loro probabilità condizionale di occorrere dato "z". Lo stesso vale per i suffissi *-tore*, *-ico*.

5.2 Una riflessione teorica

L'evidenza sperimentale proposta offre un interessante spunto di rianalisi teorica. Gli effetti algoritmici di emergenza della percezione della struttura interna delle parole, attraverso la dinamica di predizione crescente delle sotto-stringhe condivise da più forme, sembrano validare approcci teorici sullo stampo del *Word-and-Paradigm* (James P Blevins 2006; Matthews 1991), per cui una parola morfologicamente complessa è percepita come scomponibile in sotto-costituenti nella misura in cui questi sotto-costituenti sono condivisi all'interno di una rete di altre forme dello stesso paradigma o famiglia lessicale. Le relazioni tra forme non sarebbero pertanto il risultato dell'applicazione di regole produttive, piuttosto il risultato dell'integrazione dinamica di forme all'interno di uno stesso paradigma o serie derivazionale (Hathout 2009), come nel caso in esame.

La nozione stessa di morfema acquisisce una identità del tutto funzionale, anziché formale, caratterizzata dal suo ruolo in chiave elaborativa, pronta a passare in secondo piano nel caso di specifiche condizioni di input che tendano ad offuscare la percezione della struttura morfologica e delle possibili relazioni con altre forme (come nel caso delle forme ad alta frequenza nell'Esperimento 2).

L'emergenza *processing*-dipendente dei morfemi risulta anche dipendente dalle dinamiche di memorizzazione competitiva tra forme. Forme ad alta frequenza (si veda Esperimento 2) tenderanno ad essere memorizzate come un tutto, e apprese tendenzialmente in isolamento,

proprio in funzione della alta frequenza con la quale occorrono nell'input della rete - e del parlante. Il radicamento di queste tracce di memoria porta ad offuscare la percezione delle possibili relazioni superficiali, riducendo, pertanto, la percezione della loro struttura morfologica.

A ciò si contrappone l'effetto strutturale che maggiormente emerge negli Esperimenti 1 e 3. Qui, nella condizione di input con frequenze bilanciate, ciò che maggiormente emerge è la sensibilità alla struttura morfologica, ovvero quelle sequenze morfemiche condivise da più forme che emergono durante le epoche di apprendimento come unità fondamentali di elaborazione strutturale (Marzi 2020). Senza informazioni di carattere morfo-sintattico, le strutture lessicali emergono proprio dalla ridondanza formale nei dati di input.

5.3 Possibili sviluppi

Va evidenziato, naturalmente, che l'impianto sperimentale condotto con un tipo di rete neurale con un solo livello di rappresentazione non permette di mettere a fattore la dimensione semantica, ed in particolare l'interazione tra l'accesso al contenuto semantico di una parola e la sua elaborazione più superficiale.

Senza dubbio, un'architettura a più livelli, in grado di rappresentare anche le relazioni semantiche tra - ad esempio - una forma derivata e la sua base, o tra forme morfologicamente complesse che condividono tra loro la stessa base, offrirebbe un interessante evidenza aggiuntiva, ponendo in termini algoritmici la relazione tra forma e significato, e i suoi effetti sulle dinamiche di elaborazione.

Va enfatizzato, però, che nel tentativo di modellare i meccanismi di base che governano la memorizzazione e l'elaborazione delle parole, compatibilmente con l'evidenza comportamentale psico-cognitiva e i vincoli neuro-funzionali esplicitati nella parte teorica, la TSOM (*Temporal Self-Organising Map*) si è rivelata estremamente efficace nel mettere in luce gli effetti distribuzionali sulla percezione della struttura morfologica dell'input, e sulle dinamiche di apprendimento.

Resta palese che pur decidendo di focalizzarsi su una specifica manifestazione della lingua, ovvero il fenomeno morfologico della derivazione, è impossibile pensare di vederla totalmente fuori dal suo contesto.

Bibliografia

- Akmajian, A. (1979). "Linguistics: An Introduction to Language and Communication". In:
Baayen, Harald (1992). "Quantitative aspects of morphological productivity". In: *Yearbook of morphology 1991*. Springer, pp. 109–149.
- Balota, David A, Giovanni B Flores d'Arcais e Keith Rayner (1990). *Comprehension processes in reading*. Routledge.
- Bertinetto, Pier Marco et al. (2005). "Colfis (corpus e lessico di frequenza dell'italiano scritto)". In: Available on <http://www.istc.cnr.it/material/database>, pp. 67–73.
- Blevins, James P (2006). "Word-based morphology". In: *Journal of linguistics* 42.3, pp. 531–573.
— (2016). *Word and Paradigm Morphology*. Oxford: Oxford University Press.
- Booij, G. et al. (2000). "Morphology :an international handbook on inflection and word-formation."
In:
- Bortolini, Umberta, Carlo Tagliavini e Antonio Zampolli (1971). *Lessico di frequenza della lingua italiana contemporanea*. IBM Italia.
- Bruner, J.S. e R. Watson (1983). *Child's Talk: Learning to Use Language*. W.W. Norton.
- Burani, Cristina e Alfonso Caramazza (1987). "Representation and processing of derived words". In: *Language and cognitive processes* 2.3-4, pp. 217–227.
- Burani, Cristina e Anna M Thornton (2011). "The interplay of root, suffix and whole-word frequency in processing derived words". In: *Morphological structure in language processing*. De Gruyter Mouton, pp. 157–208.
- Burzio, Luigi (2004). "Paradigmatic and syntagmatic relations in Italian verbal inflection". In: *Amsterdam studies in the theory and history of linguistic science series 4* 258, pp. 17–44.
- Bybee, Joan (2010). *Language, usage and cognition*. Cambridge University Press.

- Bybee, Joan L (1985). *Morphology: A study of the relation between meaning and form*. Vol. 9. John Benjamins Publishing.
- (1995). “Regular morphology and the lexicon.” In: *Language and Cognitive Processes* 10, pp. 425–455.
- Cacciari, Cristina (2011). *Psicologia del linguaggio*. Il mulino.
- Caramazza, A., A. Laudanna e C. Romani (1988). “Lexical access and inflectional morphology”. In: *Cognition* 28, pp. 297–332.
- Catani, Marco, Derek K Jones e Dominic H Ffytche (2005). “Perisylvian language networks of the human brain”. In: *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* 57.1, pp. 8–16.
- Chomsky, Noam (1957). *Syntactic structures*.
- (1959). “Review of B.F. Skinner, Verbal Behavior”. In: *Language* 35, pp. 26–58.
- Churchland, Patricia Smith (1989). *Neurophilosophy: Toward a unified science of the mind-brain*. MIT press.
- Clahsen, Harald (dic. 2006). “Dual-Mechanism Morphology”. In: *Encyclopedia of language and linguistics* 4. DOI: 10.1016/B0-08-044854-2/04252-8.
- Dressler, Wolfgang U e Mária Ladányi (2000). “Productivity in word formation (WF): a morphological approach”. In: *Acta Linguistica Hungarica* 47.1, pp. 103–145.
- Elman, Jeffrey L (2004). “An alternative view of the mental lexicon”. In: *Trends in cognitive sciences* 8.7, pp. 301–306.
- Emmorey, Karen D e Victoria A Fromkin (1988). “The mental lexicon”. In: *Linguistics: the Cambridge survey* 3, pp. 124–149.
- Ferro, Marcello, Claudia Marzi e Vito Pirrelli (2011). “A self-organizing model of word storage and processing: implications for morphology learning”. In: *Lingue e linguaggio* 10.2, pp. 209–226.
- Ferro, Marcello, Giovanni Pezzulo e Vito Pirrelli (2010). “Morphology, memory and the mental lexicon”. In: *Lingue e Linguaggio* 9.2, pp. 199–.
- Floreato, Dario e Claudio Mattiussi (2002). *Manuale sulle reti neurali*. Rapp. tecn. Il mulino.
- Fodor, Jerry A (1990). “A theory of content and other essays”. In:

- Fodor J., Pylyshyn Z. (1988). "Connectionism and cognitive architecture: A critical analysis". In: *Cognition* 28, pp. 3–71.
- Forster, Kenneth I e Elizabeth S Bednall (1976). "Terminating and exhaustive search in lexical access". In: *Memory & Cognition* 4.1, pp. 53–61.
- Frege, Gottlob (1948). "Sense and reference". In: *The philosophical review* 57.3, pp. 209–230.
- Gaeta, Livio e Davide Ricca (2003). "Frequency and productivity in Italian derivation: A comparison between corpus-based and lexicographical data". In:
- Garman, Michael (1992). "Psycholinguistics". In: *Child Language Teaching and Therapy* 8.1, pp. 102–104.
- Gazzaniga, Michael S et al. (2015). *Neuroscienze cognitive*. Zanichelli.
- Giraud, H el ene e Serena Dal Maso (2016). "Suffix perceptual salience in morphological masked priming". In: *Lingue e Linguaggio* 15.1, pp. 85–106.
- Gonnerman, Laura M, Mark S Seidenberg e Elaine S Andersen (2007). "Graded semantic and phonological similarity effects in priming: Evidence for a distributed connectionist approach to morphology." In: *Journal of experimental psychology: General* 136.2, p. 323.
- Hathout, Nabil (2009). "Acquisition of morphological families and derivational series from a machine readable dictionary". In: *arXiv preprint arXiv:0905.1609*.
- Hay, Jennifer (2001). "Lexical frequency in morphology: Is everything relative?" In:
- Hebb, Donald Olding (1949). *The organisation of behaviour: a neuropsychological theory*. Science Editions New York.
- Jackendoff, R. (2002). "Foundations of Language: Brain, Meaning, Grammar, Evolution". In: *Language in Society* 32.3, pp. 415–418.
- Kohonen, Teuvo (2001). "Learning vector quantization". In: *Self-organizing maps*. Springer, pp. 245–261.
- Laudanna, Alessandro e Cristina Burani (1995). "Distributional properties of derivational affixes: Implications for processing". In: *Morphological aspects of language processing*, pp. 345–364.
- Laudanna, Alessandro, Anna Maria Thornton et al. (1995). "Un corpus dell'italiano scritto contemporaneo dalla parte del ricevente". In: *III giornate internazionali di analisi statistica dei dati testuali* 1, pp. 103–109.

- Lenci, Alessandro, Simonetta Montemagni e Vito Pirrelli (2005). *Testo e computer. Introduzione alla linguistica computazionale*. Carocci editore.
- Libben, G. (2005). “Everything is Psycholinguistics: Material and Methodological Considerations in the Study of Compound Processing”. In: *The Canadian Journal of Linguistics / La revue canadienne de linguistique* 50, pp. 267–283.
- Libben, Gary (1998). “Semantic transparency in the processing of compounds: Consequences for representation, processing, and impairment”. In: *Brain and language* 61.1, pp. 30–44.
- Lieber, Rochelle (2017). “Derivational morphology”. In: *Oxford Research Encyclopedia of Linguistics*.
- Markman, Arthur B. e Eric Dietrich (2000). “Extending the classical view of representation”. In: *Trends in Cognitive Sciences* 4.12, pp. 470–475.
- Marr, David e Tomaso Poggio (1979). “A computational theory of human stereo vision”. In: *Proceedings of the Royal Society of London. Series B. Biological Sciences* 204.1156, pp. 301–328.
- Marslen-Wilson, W. et al. (1994). “Morphology and meaning in the English mental lexicon.” In: *Psychological Review* 101, pp. 3–33.
- Marzi, Claudia (2020). “Modeling Word Learning and Processing with Recurrent Neural Networks”. In: *Information* 11.6, p. 320.
- Marzi, Claudia, Marcello Ferro, Franco Alberto Cardillo et al. (2016). “Effects of frequency and regularity in an integrative model of word storage and processing”. In: *Italian Journal of Linguistics* 28.1, pp. 79–114.
- Marzi, Claudia, Marcello Ferro, Ouafae Nahli et al. (2018). “Evaluating inflectional complexity crosslinguistically: A processing perspective”. In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.
- Marzi, Claudia, Marcello Ferro e Vito Pirrelli (2014). “Morphological structure through lexical parsability”. In: *Lingue e linguaggio* 13.2, pp. 263–.
- (2019). “A Processing-Oriented Investigation of Inflectional Complexity”. In: *Frontiers in Communication* 4.
- Matthews, Peter H (1991). “Morphology (cambridge textbooks in linguistics)”. In: *New York: Cambridge University*.

- McClelland, James L, David E Rumelhart, PDP Research Group et al. (1986). *Parallel distributed processing*. Vol. 2. MIT press Cambridge, MA.
- Miller, George A (1967). "The psychology of communication". In: *Human Resource Management* 6.3, p. 43.
- Morton, J. (1969). "Interaction of information in word recognition." In: *Psychological Review* 76, pp. 165–178.
- (1979). "Facilitation in word recognition: Experiments causing change in the logogen model". In: *Processing of visible language*. Springer, pp. 259–268.
- (1980). "The logogen model and orthographic structure". In: *Cognitive processes in spelling*.
- Moscoso del Prado Martin, Fermin et al. (2004). "Morphological family size in a morphologically rich language: the case of Finnish compared with Dutch and Hebrew." In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30.6, p. 1271.
- Oldfield, R. C. (1966). "Things, Words and the Brain*". In: *Quarterly Journal of Experimental Psychology* 18, pp. 340–353.
- Osgood, C. e T. Sebeok (1954). "Psycholinguistics: a survey of theory and research problems." In: *Journal of abnormal psychology* 49 4, Part 2, pp. 1–203.
- Pinker, S. e Alan S. Prince (1988). "On language and connectionism: Analysis of a parallel distributed processing model of language acquisition". In: *Cognition* 28, pp. 73–193.
- Pinker, Steven (1994). "The Language Instinct: How the Mind Creates". In: *Language*. New York: Harper Collins.
- Pinker e Ullman (2002). "The Past and Future of the Past Tense Debate". In: *Trends in cognitive sciences* 6.11, pp. 456–463.
- Pirrelli, Vito (2007). "Psycho-Computational Issues in Morphology Learning and Processing: An Overture". In: *Lingue e Linguaggio* 2.
- Pirrelli, Vito e Marco Battista (2000). "The paradigmatic dimension of stem allomorphy in Italian verb inflection: 2628". In: *The Italian Journal of Linguistics* 12, p. 307.
- Pirrelli, Vito, Marcello Ferro e Claudia Marzi (2015). "Computational complexity of abstractive morphology". In: *Understanding and measuring morphological complexity*, pp. 141–166.

- Plag, Ingo e Laura Winther Balling (2020). “Derivational morphology: An integrative perspective on some fundamental questions”. In: *Word Knowledge and Word Usage*. De Gruyter Mouton, pp. 295–335.
- Pullum, Geoffrey K e Barbara C Scholz (2002). “Empirical assessment of stimulus poverty arguments”. In: *The linguistic review* 19.1-2, pp. 9–50.
- Putnam, Hilary (1987). *The Many Faces of Realism*. Open Court.
- (1993). “The Question of Realism”. In: *Words and Life*. A cura di James Conant. Harvard University Press, pp. 295–312.
- Rizzi, L. (1994). *La conoscenza del linguaggio, in SISSA (a cura di), Neuroscienze e scienze cognitive*. Napoli: CUEN.
- Rueckl, Jay G e Bruno Galantucci (2005). “morphological priming”. In:
- Rumelhart, David E. e James L. McClelland (1987). “On Learning the Past Tenses of English Verbs”. In: *Parallel Distributed Processing. Explorations in the Microstructures of Cognition*. A cura di David E. Rumelhart, James L. McClelland e the PDP Research Group. Vol. 2 Psychological and Biological Models. MIT Press, pp. 216–271.
- Schreuder, Robert e R Harald Baayen (1995). “Modeling morphological processing”. In: *Morphological aspects of language processing* 2, pp. 257–294.
- Seidenberg, Mark S (1995). “Visual word recognition: An overview.” In:
- Seidenberg, Mark S. e James L. McClelland (1989). “A distributed, developmental model of word recognition and naming.” In: *Psychological review* 96 4, pp. 523–68.
- Singleton, David (1999). *Exploring the Second Language Mental Lexicon*. Cambridge Applied Linguistics. Cambridge University Press.
- Skinner, Burrhus Frederic (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.
- Spencer, Andrew (1991). *Morphological theory: An introduction to word structure in generative grammar*. Wiley-Blackwell.
- Spencer, Andrew e Arnold M Zwicky (2001). *The handbook of morphology*. Wiley-Blackwell.
- Taft, Marcus e Kenneth I Forster (1975). “Lexical storage and retrieval of prefixed words”. In: *Journal of verbal learning and verbal behavior* 14.6, pp. 638–647.

- Talamo, Luigi e Chiara Celata (2011). “Toward a morphological analysis of the Italian lexicon: developing tools for a corpus-based approach”. In: *Quaderni del Laboratorio di Linguistica* 1.10.
- Talamo, Luigi, Chiara Celata e Pier Marco Bertinetto (2016). “DerIvaTario: An annotated lexicon of Italian derivatives”. In: *Word Structure* 9.1, pp. 72–102.
- (apr. 2016). “DerIvaTario: An annotated lexicon of Italian derivatives”. In: *Word Structure* 9, pp. 72–102. DOI: 10.3366/word.2016.0087.
- Taylor, K. e J. Fodor (1990). “Psychosemantics: The Problem of Meaning in the Philosophy of Mind.” In: *Noûs* 24, p. 181.
- Ulam, S. (1958). “John von Neumann 1903-1957”. In: *Bulletin of the American Mathematical Society* 64, pp. 1–49.
- Vallar, Giuseppe (2020). “The methodological foundations of neuropsychology”. In: *Handbook of clinical and experimental neuropsychology*. Psychology Press, pp. 95–132.
- Westermann, Gert e Kim Plunkett (2007). “Connectionist models of inflection processing”. In: *Lingue e Linguaggio* 6.2, pp. 291–.
- Wurm, Lee H. et al. (2006). “Dynamics of the auditory comprehension of prefixed words: Cohort entropies and conditional root uniqueness points”. In: *The Mental Lexicon* 1, pp. 125–146.
- Zeidenberg, Matthew (1990). “Neural Networks and Artificial Intelligence”. In: *Ellis Horwood, New York*.

APPENDICE

Si riportano in questa sezione l'elenco delle forme derivate dell'italiano utilizzate per gli addestramenti della rete neurale artificiale TSOM, unitamente alla codifica ortogonale dei 23 simboli formanti l'input stesso.

Per ogni forma derivata è indicata la frequenza di 5 occorrenze per il regime di addestramento uniforme (Esperimento 1), e la frequenza derivante da corpus (P. M. Bertinetto et al. 2005) (Esperimento 2).

Per quanto concerne l'esperimento con famiglie (Esperimento 3) si riportano le forme con frequenza bilanciata pari a 5 occorrenze ordinate per famiglia di appartenenza.

Segue la codifica ortogonale dei simboli, rappresentati con vettori *one-hot*.

Infine, è riportato il codice per le analisi quantitative condotte con il software *R* (<https://cran.r-project.org>).

tsom input	frequenze bilanciate	frequenze da corpus
#,A,B,I,T,A,Z,I,O,N,E,\$	5	160
#,A,B,N,E,G,A,Z,I,O,N,E,\$	5	6
#,A,B,O,L,I,Z,I,O,N,E,\$	5	39
#,A,B,R,O,G,A,Z,I,O,N,E,\$	5	6
#,A,C,C,E,L,E,R,A,Z,I,O,N,E,\$	5	29
#,A,C,C,E,N,S,I,O,N,E,\$	5	73
#,A,C,C,E,N,T,U,A,Z,I,O,N,E,\$	5	6
#,A,C,C,E,T,T,A,Z,I,O,N,E,\$	5	24
#,A,C,C,L,A,M,A,Z,I,O,N,E,\$	5	6
#,A,D,E,S,I,O,N,E,\$	5	73
#,A,F,F,E,R,M,A,Z,I,O,N,E,\$	5	104
#,A,F,F,I,L,I,A,Z,I,O,N,E,\$	5	6
#,A,G,E,V,O,L,A,Z,I,O,N,E,\$	5	25
#,A,G,G,R,E,S,S,I,O,N,E,\$	5	105
#,A,G,I,T,A,Z,I,O,N,E,\$	5	27
#,A,L,L,U,C,I,N,A,Z,I,O,N,E,\$	5	15
#,A,L,L,U,S,I,O,N,E,\$	5	26
#,A,L,T,E,R,A,Z,I,O,N,E,\$	5	31
#,A,M,B,I,Z,I,O,N,E,\$	5	76
#,A,M,M,I,R,A,Z,I,O,N,E,\$	5	46
#,A,M,M,I,S,S,I,O,N,E,\$	5	39
#,A,N,I,M,A,Z,I,O,N,E,\$	5	30
#,A,N,T,I,C,I,P,A,Z,I,O,N,E,\$	5	31
#,A,P,P,A,R,I,Z,I,O,N,E,\$	5	64
#,A,P,P,L,I,C,A,Z,I,O,N,E,\$	5	110
#,A,P,P,R,O,S,S,I,M,A,Z,I,O,N,E,\$	5	18
#,A,S,P,I,R,A,Z,I,O,N,E,\$	5	57
#,A,S,S,E,G,N,A,Z,I,O,N,E,\$	5	32
#,A,T,T,R,A,Z,I,O,N,E,\$	5	68
#,B,E,N,E,D,I,Z,I,O,N,E,\$	5	32
#,C,A,N,C,E,L,L,A,Z,I,O,N,E,\$	5	22
#,C,A,R,B,U,R,A,Z,I,O,N,E,\$	5	10
#,C,E,S,S,A,Z,I,O,N,E,\$	5	11
#,C,I,R,C,O,L,A,Z,I,O,N,E,\$	5	134
#,C,I,T,A,Z,I,O,N,E,\$	5	56
#,C,O,L,L,A,B,O,R,A,Z,I,O,N,E,\$	5	266
#,C,O,L,L,O,C,A,Z,I,O,N,E,\$	5	24
#,C,O,L,L,U,T,T,A,Z,I,O,N,E,\$	5	16
#,C,O,L,T,I,V,A,Z,I,O,N,E,\$	5	20
#,C,O,M,B,I,N,A,Z,I,O,N,E,\$	5	67
#,C,O,M,M,O,Z,I,O,N,E,\$	5	24
#,C,O,M,P,A,S,S,I,O,N,E,\$	5	11
#,C,O,M,P,E,T,I,Z,I,O,N,E,\$	5	79
#,C,O,M,U,N,I,C,A,Z,I,O,N,E,\$	5	203
#,C,O,N,D,I,V,I,S,I,O,N,E,\$	5	24
#,C,O,N,N,O,T,A,Z,I,O,N,E,\$	5	17
#,C,O,N,S,E,R,V,A,Z,I,O,N,E,\$	5	57
#,C,O,N,S,I,D,E,R,A,Z,I,O,N,E,\$	5	194

#,C,O,N,S,O,L,A,Z,I,O,N,E,\$	5	31
#,C,O,N,S,U,L,T,A,Z,I,O,N,E,\$	5	72
#,C,O,N,T,E,S,T,A,Z,I,O,N,E,\$	5	61
#,C,O,N,T,R,A,D,D,I,Z,I,O,N,E,\$	5	75
#,C,O,N,T,R,A,T,T,A,Z,I,O,N,E,\$	5	19
#,C,O,N,T,R,A,Z,I,O,N,E,\$	5	40
#,C,O,N,V,E,R,S,A,Z,I,O,N,E,\$	5	113
#,C,R,E,A,Z,I,O,N,E,\$	5	132
#,D,E,L,E,G,A,Z,I,O,N,E,\$	5	87
#,D,E,M,O,L,I,Z,I,O,N,E,\$	5	14
#,D,E,S,T,I,N,A,Z,I,O,N,E,\$	5	60
#,D,E,T,E,R,M,I,N,A,Z,I,O,N,E,\$	5	57
#,D,I,C,H,I,A,R,A,Z,I,O,N,E,\$	5	343
#,D,I,M,O,S,T,R,A,Z,I,O,N,E,\$	5	63
#,D,I,S,P,E,R,A,Z,I,O,N,E,\$	5	102
#,D,I,V,U,L,G,A,Z,I,O,N,E,\$	5	21
#,D,I,Z,I,O,N,E,\$	5	15
#,D,O,N,A,Z,I,O,N,E,\$	5	22
#,D,O,T,A,Z,I,O,N,E,\$	5	16
#,E,B,O,L,L,I,Z,I,O,N,E,\$	5	17
#,E,C,C,I,T,A,Z,I,O,N,E,\$	5	31
#,E,D,U,C,A,Z,I,O,N,E,\$	5	153
#,E,F,F,U,S,I,O,N,E,\$	5	9
#,E,L,A,B,O,R,A,Z,I,O,N,E,\$	5	42
#,E,L,I,M,I,N,A,Z,I,O,N,E,\$	5	50
#,E,M,A,N,A,Z,I,O,N,E,\$	5	10
#,E,M,A,N,C,I,P,A,Z,I,O,N,E,\$	5	12
#,E,M,A,R,G,I,N,A,Z,I,O,N,E,\$	5	29
#,E,M,I,G,R,A,Z,I,O,N,E,\$	5	19
#,E,R,O,G,A,Z,I,O,N,E,\$	5	21
#,E,S,C,L,A,M,A,Z,I,O,N,E,\$	5	12
#,E,S,I,B,I,Z,I,O,N,E,\$	5	57
#,E,S,I,T,A,Z,I,O,N,E,\$	5	37
#,E,S,P,L,O,R,A,Z,I,O,N,E,\$	5	30
#,F,A,B,B,R,I,C,A,Z,I,O,N,E,\$	5	28
#,F,E,D,E,R,A,Z,I,O,N,E,\$	5	202
#,F,O,N,D,A,Z,I,O,N,E,\$	5	131
#,F,O,R,M,A,Z,I,O,N,E,\$	5	310
#,G,E,N,E,R,A,Z,I,O,N,E,\$	5	244
#,G,U,A,R,I,G,I,O,N,E,\$	5	31
#,I,L,L,U,S,I,O,N,E,\$	5	121
#,I,L,L,U,S,T,R,A,Z,I,O,N,E,\$	5	27
#,I,M,I,T,A,Z,I,O,N,E,\$	5	38
#,I,M,M,A,G,I,N,A,Z,I,O,N,E,\$	5	51
#,I,M,M,E,R,S,I,O,N,E,\$	5	12
#,I,M,M,I,S,S,I,O,N,E,\$	5	12
#,I,N,D,I,C,A,Z,I,O,N,E,\$	5	162
#,I,N,F,O,R,M,A,Z,I,O,N,E,\$	5	439
#,I,N,I,B,I,Z,I,O,N,E,\$	5	11
#,I,N,I,Z,I,A,Z,I,O,N,E,\$	5	30

#,I,N,T,E,R,P,R,E,T,A,Z,I,O,N,E,\$	5	154
#,I,N,T,U,I,Z,I,O,N,E,\$	5	53
#,I,R,R,I,T,A,Z,I,O,N,E,\$	5	25
#,I,S,P,I,R,A,Z,I,O,N,E,\$	5	62
#,L,A,V,O,R,A,Z,I,O,N,E,\$	5	54
#,L,I,M,I,T,A,Z,I,O,N,E,\$	5	41
#,L,I,Q,U,I,D,A,Z,I,O,N,E,\$	5	38
#,M,E,D,I,T,A,Z,I,O,N,E,\$	5	44
#,M,O,B,I,L,I,T,A,Z,I,O,N,E,\$	5	22
#,M,O,D,E,R,A,Z,I,O,N,E,\$	5	25
#,M,U,T,A,Z,I,O,N,E,\$	5	22
#,M,U,T,I,L,A,Z,I,O,N,E,\$	5	16
#,N,A,R,R,A,Z,I,O,N,E,\$	5	20
#,N,A,V,I,G,A,Z,I,O,N,E,\$	5	37
#,N,U,T,R,I,Z,I,O,N,E,\$	5	15
#,O,B,B,L,I,G,A,Z,I,O,N,E,\$	5	31
#,O,C,C,U,P,A,Z,I,O,N,E,\$	5	164
#,O,P,E,R,A,Z,I,O,N,E,\$	5	535
#,O,S,C,I,L,L,A,Z,I,O,N,E,\$	5	22
#,O,S,S,E,R,V,A,Z,I,O,N,E,\$	5	125
#,P,A,R,T,E,C,I,P,A,Z,I,O,N,E,\$	5	163
#,P,E,N,E,T,R,A,Z,I,O,N,E,\$	5	13
#,P,E,R,C,U,S,S,I,O,N,E,\$	5	7
#,P,E,R,Q,U,I,S,I,Z,I,O,N,E,\$	5	48
#,P,I,A,N,T,A,G,I,O,N,E,\$	5	5
#,P,R,E,N,O,T,A,Z,I,O,N,E,\$	5	24
#,P,R,E,O,C,C,U,P,A,Z,I,O,N,E,\$	5	163
#,P,R,E,P,A,R,A,Z,I,O,N,E,\$	5	174
#,P,R,E,S,E,N,T,A,Z,I,O,N,E,\$	5	110
#,P,R,E,S,T,A,Z,I,O,N,E,\$	5	131
#,P,R,E,V,I,S,I,O,N,E,\$	5	156
#,P,R,I,V,A,T,I,Z,Z,A,Z,I,O,N,E,\$	5	129
#,P,R,O,D,U,Z,I,O,N,E,\$	5	364
#,P,R,O,G,E,T,T,A,Z,I,O,N,E,\$	5	21
#,P,R,O,G,R,A,M,M,A,Z,I,O,N,E,\$	5	43
#,P,R,O,I,E,Z,I,O,N,E,\$	5	79
#,P,R,O,L,I,F,E,R,A,Z,I,O,N,E,\$	5	18
#,P,R,O,M,O,Z,I,O,N,E,\$	5	88
#,P,R,O,T,E,Z,I,O,N,E,\$	5	147
#,P,R,O,V,O,C,A,Z,I,O,N,E,\$	5	70
#,P,U,B,B,L,I,C,A,Z,I,O,N,E,\$	5	81
#,P,U,N,I,Z,I,O,N,E,\$	5	76
#,Q,U,O,T,A,Z,I,O,N,E,\$	5	43
#,R,A,C,C,O,M,A,N,D,A,Z,I,O,N,E,\$	5	49
#,R,A,D,I,A,Z,I,O,N,E,\$	5	43
#,R,A,P,P,R,E,S,E,N,T,A,Z,I,O,N,E,\$	5	85
#,R,E,A,Z,I,O,N,E,\$	5	296
#,R,E,C,E,S,S,I,O,N,E,\$	5	77
#,R,E,C,L,U,S,I,O,N,E,\$	5	24
#,R,E,D,A,Z,I,O,N,E,\$	5	89

#,R,E,G,I,S,T,R,A,Z,I,O,N,E,\$	5	68
#,R,E,L,A,Z,I,O,N,E,\$	5	431
#,R,E,T,R,I,B,U,Z,I,O,N,E,\$	5	59
#,R,E,V,I,S,I,O,N,E,\$	5	58
#,R,I,D,U,Z,I,O,N,E,\$	5	134
#,R,I,F,L,E,S,S,I,O,N,E,\$	5	122
#,R,I,F,O,N,D,A,Z,I,O,N,E,\$	5	107
#,R,I,P,A,R,A,Z,I,O,N,E,\$	5	52
#,R,I,V,E,L,A,Z,I,O,N,E,\$	5	124
#,S,A,N,Z,I,O,N,E,\$	5	101
#,S,E,G,N,A,L,A,Z,I,O,N,E,\$	5	64
#,S,E,P,A,R,A,Z,I,O,N,E,\$	5	115
#,S,I,T,U,A,Z,I,O,N,E,\$	5	999
#,S,O,D,D,I,S,F,A,Z,I,O,N,E,\$	5	177
#,S,O,S,P,E,N,S,I,O,N,E,\$	5	101
#,S,P,E,C,U,L,A,Z,I,O,N,E,\$	5	53
#,S,P,E,D,I,Z,I,O,N,E,\$	5	62
#,S,P,I,E,G,A,Z,I,O,N,E,\$	5	188
#,S,T,A,Z,I,O,N,E,\$	5	224
#,S,U,C,C,E,S,S,I,O,N,E,\$	5	45
#,T,E,N,T,A,Z,I,O,N,E,\$	5	105
#,T,R,A,D,I,Z,I,O,N,E,\$	5	329
#,T,R,A,D,U,Z,I,O,N,E,\$	5	67
#,T,R,A,N,S,I,Z,I,O,N,E,\$	5	45
#,T,R,A,S,F,O,R,M,A,Z,I,O,N,E,\$	5	108
#,T,R,A,S,M,I,S,S,I,O,N,E,\$	5	301
#,T,R,A,S,P,O,S,I,Z,I,O,N,E,\$	5	8
#,U,C,C,I,S,I,O,N,E,\$	5	47
#,U,M,I,L,I,A,Z,I,O,N,E,\$	5	29
#,V,A,L,U,T,A,Z,I,O,N,E,\$	5	107
#,V,A,R,I,A,Z,I,O,N,E,\$	5	85
#,V,E,N,E,R,A,Z,I,O,N,E,\$	5	8
#,V,E,N,T,I,L,A,Z,I,O,N,E,\$	5	7
#,V,I,B,R,A,Z,I,O,N,E,\$	5	42
#,V,I,O,L,A,Z,I,O,N,E,\$	5	85
#,V,I,S,I,O,N,E,\$	5	173
#,V,O,C,A,Z,I,O,N,E,\$	5	75
#,V,O,T,A,Z,I,O,N,E,\$	5	51
#,A,M,M,I,N,I,S,T,R,A,Z,I,O,N,E,\$	5	338
#,A,G,G,R,E,G,A,Z,I,O,N,E,\$	5	30
#,C,O,L,O,N,I,Z,Z,A,Z,I,O,N,E,\$	5	10
#,D,E,T,E,N,Z,I,O,N,E,\$	5	28
#,D,I,S,P,O,S,I,Z,I,O,N,E,\$	5	496
#,I,N,T,E,R,R,O,G,A,Z,I,O,N,E,\$	5	32
#,I,N,T,R,O,D,U,Z,I,O,N,E,\$	5	57
#,C,O,N,C,E,Z,I,O,N,E,\$	5	59
#,I,M,P,R,E,S,S,I,O,N,E,\$	5	196
#,M,E,D,I,A,Z,I,O,N,E,\$	5	50
#,M,O,T,I,V,A,Z,I,O,N,E,\$	5	98
#,O,B,I,E,Z,I,O,N,E,\$	5	55

#,C,O,N,D,U,Z,I,O,N,E,\$	5	24
#,G,E,S,T,I,O,N,E,\$	5	224
#,A,C,C,E,L,E,R,A,T,O,R,E,\$	5	23
#,A,C,C,O,M,P,A,G,N,A,T,O,R,E,\$	5	14
#,A,C,C,U,S,A,T,O,R,E,\$	5	17
#,A,D,O,R,A,T,O,R,E,\$	5	7
#,A,G,G,R,E,S,S,O,R,E,\$	5	37
#,A,L,B,E,R,G,A,T,O,R,E,\$	5	28
#,A,L,I,M,E,N,T,A,T,O,R,E,\$	5	8
#,A,L,L,E,N,A,T,O,R,E,\$	5	177
#,A,M,M,I,N,I,S,T,R,A,T,O,R,E,\$	5	230
#,A,M,M,I,R,A,T,O,R,E,\$	5	29
#,A,M,M,O,N,I,T,O,R,E,\$	5	2
#,A,N,I,M,A,T,O,R,E,\$	5	10
#,A,S,C,E,N,S,O,R,E,\$	5	55
#,A,S,C,O,L,T,A,T,O,R,E,\$	5	30
#,A,S,S,A,G,G,I,A,T,O,R,E,\$	5	2
#,A,S,S,A,L,I,T,O,R,E,\$	5	6
#,A,S,S,A,L,T,A,T,O,R,E,\$	5	2
#,A,S,S,E,R,T,O,R,E,\$	5	6
#,A,S,S,I,C,U,R,A,T,O,R,E,\$	5	9
#,B,O,B,I,N,A,T,O,R,E,\$	5	7
#,B,O,L,L,I,T,O,R,E,\$	5	4
#,C,A,C,C,I,A,T,O,R,E,\$	5	44
#,C,A,L,C,I,A,T,O,R,E,\$	5	109
#,C,A,L,C,O,L,A,T,O,R,E,\$	5	18
#,C,A,R,B,U,R,A,T,O,R,E,\$	5	7
#,C,A,R,I,C,A,T,O,R,E,\$	5	6
#,C,L,I,M,A,T,I,Z,Z,A,T,O,R,E,\$	5	8
#,C,O,L,L,A,B,O,R,A,T,O,R,E,\$	5	137
#,C,O,L,T,I,V,A,T,O,R,E,\$	5	11
#,C,O,M,M,E,N,T,A,T,O,R,E,\$	5	28
#,C,O,M,P,E,T,I,T,O,R,E,\$	5	2
#,C,O,M,P,O,S,I,T,O,R,E,\$	5	65
#,C,O,M,P,R,A,T,O,R,E,\$	5	28
#,C,O,N,D,U,T,T,O,R,E,\$	5	56
#,C,O,N,F,E,S,S,A,T,O,R,E,\$	5	7
#,C,O,N,O,S,C,I,T,O,R,E,\$	5	13
#,C,O,N,Q,U,I,S,T,A,T,O,R,E,\$	5	6
#,C,O,N,S,E,R,V,A,T,O,R,E,\$	5	35
#,C,O,N,S,U,M,A,T,O,R,E,\$	5	139
#,C,O,N,T,E,N,I,T,O,R,E,\$	5	67
#,C,O,N,T,E,S,T,A,T,O,R,E,\$	5	9
#,C,O,N,T,R,O,L,L,O,R,E,\$	5	13
#,C,O,O,R,D,I,N,A,T,O,R,E,\$	5	38
#,C,O,R,R,I,D,O,R,E,\$	5	62
#,C,O,R,R,U,T,T,O,R,E,\$	5	11
#,C,O,S,T,R,U,T,T,O,R,E,\$	5	93
#,C,R,E,A,T,O,R,E,\$	5	38
#,C,R,E,D,I,T,O,R,E,\$	5	14

#,C,U,L,T,O,R,E,\$	5	18
#,C,U,R,A,T,O,R,E,\$	5	13
#,D,A,T,O,R,E,\$	5	34
#,D,E,N,O,M,I,N,A,T,O,R,E,\$	5	12
#,D,E,T,O,N,A,T,O,R,E,\$	5	5
#,D,E,T,R,A,T,T,O,R,E,\$	5	11
#,D,I,F,E,N,S,O,R,E,\$	5	130
#,D,I,F,F,U,S,O,R,E,\$	5	11
#,D,I,R,E,T,T,O,R,E,\$	5	739
#,D,I,S,E,G,N,A,T,O,R,E,\$	5	21
#,D,I,S,T,R,I,B,U,T,O,R,E,\$	5	35
#,D,I,T,T,A,T,O,R,E,\$	5	35
#,D,I,V,I,N,A,T,O,R,E,\$	5	34
#,D,I,V,U,L,G,A,T,O,R,E,\$	5	10
#,D,O,M,I,N,A,T,O,R,E,\$	5	12
#,D,O,N,A,T,O,R,E,\$	5	25
#,E,D,U,C,A,T,O,R,E,\$	5	13
#,E,L,A,B,O,R,A,T,O,R,E,\$	5	8
#,E,L,E,T,T,O,R,E,\$	5	138
#,E,S,E,C,U,T,O,R,E,\$	5	24
#,E,S,P,L,O,R,A,T,O,R,E,\$	5	14
#,E,S,P,O,S,I,T,O,R,E,\$	5	9
#,E,S,T,I,N,T,O,R,E,\$	5	9
#,E,V,A,S,O,R,E,\$	5	14
#,F,A,C,I,T,O,R,E,\$	5	179
#,F,O,N,D,A,T,O,R,E,\$	5	59
#,F,O,R,N,I,T,O,R,E,\$	5	25
#,F,U,M,A,T,O,R,E,\$	5	26
#,G,E,N,E,R,A,T,O,R,E,\$	5	5
#,G,E,S,T,O,R,E,\$	5	41
#,G,I,O,C,A,T,O,R,E,\$	5	392
#,G,O,V,E,R,N,A,T,O,R,E,\$	5	106
#,G,U,A,R,I,T,O,R,E,\$	5	7
#,G,U,I,D,A,T,O,R,E,\$	5	23
#,I,D,E,A,T,O,R,E,\$	5	22
#,I,M,I,T,A,T,O,R,E,\$	5	8
#,I,M,P,E,R,A,T,O,R,E,\$	5	89
#,I,N,C,I,S,O,R,E,\$	5	6
#,I,N,D,I,C,A,T,O,R,E,\$	5	16
#,I,N,F,O,R,M,A,T,O,R,E,\$	5	13
#,I,N,N,O,V,A,T,O,R,E,\$	5	7
#,I,N,Q,U,I,S,I,T,O,R,E,\$	5	13
#,I,N,T,E,G,R,A,T,O,R,E,\$	5	6
#,I,N,V,A,S,O,R,E,\$	5	19
#,I,N,V,E,N,T,O,R,E,\$	5	48
#,I,N,V,E,S,T,I,G,A,T,O,R,E,\$	5	162
#,I,N,V,E,S,T,I,T,O,R,E,\$	5	48
#,I,S,P,I,R,A,T,O,R,E,\$	5	16
#,I,S,T,R,U,T,T,O,R,E,\$	5	36
#,L,A,V,O,R,A,T,O,R,E,\$	5	232

#,L,E,T,T,O,R,E,\$	5	243
#,L,I,B,E,R,A,T,O,R,E,\$	5	7
#,L,I,Q,U,I,D,A,T,O,R,E,\$	5	5
#,L,O,T,T,A,T,O,R,E,\$	5	5
#,M,A,N,G,I,A,T,O,R,E,\$	5	15
#,M,E,D,I,A,T,O,R,E,\$	5	32
#,M,O,T,O,R,E,\$	5	240
#,M,U,R,A,T,O,R,E,\$	5	47
#,N,A,R,R,A,T,O,R,E,\$	5	12
#,N,A,V,I,G,A,T,O,R,E,\$	5	13
#,N,E,G,O,Z,I,A,T,O,R,E,\$	5	11
#,O,B,I,E,T,T,O,R,E,\$	5	6
#,O,P,E,R,A,T,O,R,E,\$	5	143
#,O,P,P,O,S,I,T,O,R,E,\$	5	35
#,O,P,P,R,E,S,S,O,R,E,\$	5	4
#,O,S,S,E,R,V,A,T,O,R,E,\$	5	74
#,O,T,T,U,R,A,T,O,R,E,\$	5	9
#,P,E,R,S,E,C,U,T,O,R,E,\$	5	6
#,P,E,S,C,A,T,O,R,E,\$	5	85
#,P,O,R,T,A,T,O,R,E,\$	5	32
#,P,O,S,S,E,S,S,O,R,E,\$	5	14
#,P,R,E,C,U,R,S,O,R,E,\$	5	17
#,P,R,E,D,A,T,O,R,E,\$	5	24
#,P,R,E,D,I,C,A,T,O,R,E,\$	5	12
#,P,R,E,P,A,R,A,T,O,R,E,\$	5	9
#,P,R,E,S,E,N,T,A,T,O,R,E,\$	5	29
#,P,R,O,C,U,R,A,T,O,R,E,\$	5	258
#,P,R,O,D,U,T,T,O,R,E,\$	5	20
#,P,R,O,F,E,S,S,O,R,E,\$	5	220
#,P,R,O,I,E,T,T,O,R,E,\$	5	11
#,P,R,O,M,O,T,O,R,E,\$	5	37
#,P,R,O,T,E,T,T,O,R,E,\$	5	13
#,P,R,O,V,O,C,A,T,O,R,E,\$	5	9
#,P,R,O,V,V,E,D,I,T,O,R,E,\$	5	24
#,R,A,C,C,O,G,L,I,T,O,R,E,\$	5	4
#,R,A,P,I,N,A,T,O,R,E,\$	5	93
#,R,A,P,I,T,O,R,E,\$	5	9
#,R,E,A,T,T,O,R,E,\$	5	5
#,R,E,C,E,T,T,O,R,E,\$	5	6
#,R,E,D,A,T,T,O,R,E,\$	5	27
#,R,E,D,E,N,T,O,R,E,\$	5	3
#,R,E,G,I,S,T,R,A,T,O,R,E,\$	5	19
#,R,E,L,A,T,O,R,E,\$	5	21
#,R,E,V,I,S,O,R,E,\$	5	11
#,R,I,C,E,R,C,A,T,O,R,E,\$	5	117
#,R,I,C,E,V,I,T,O,R,E,\$	5	12
#,R,I,F,O,R,M,A,T,O,R,E,\$	5	14
#,R,I,L,E,V,A,T,O,R,E,\$	5	6
#,R,I,S,P,A,R,M,I,A,T,O,R,E,\$	5	32
#,R,I,T,A,R,D,A,T,O,R,E,\$	5	15

#,R,O,D,I,T,O,R,E,\$	5	8
#,S,C,A,L,A,T,O,R,E,\$	5	23
#,S,C,E,N,E,G,G,I,A,T,O,R,E,\$	5	23
#,S,C,I,A,T,O,R,E,\$	5	13
#,S,C,R,I,T,T,O,R,E,\$	5	277
#,S,C,U,L,T,O,R,E,\$	5	25
#,S,E,D,U,T,T,O,R,E,\$	5	12
#,S,E,N,A,T,O,R,E,\$	5	216
#,S,E,R,V,I,T,O,R,E,\$	5	22
#,S,F,R,U,T,T,A,T,O,R,E,\$	5	9
#,S,I,N,T,E,T,I,Z,Z,A,T,O,R,E,\$	5	2
#,S,O,C,C,O,R,R,I,T,O,R,E,\$	5	17
#,S,O,G,N,A,T,O,R,E,\$	5	7
#,S,O,S,T,E,N,I,T,O,R,E,\$	5	57
#,S,O,T,T,O,S,C,R,I,T,T,O,R,E,\$	5	6
#,S,P,A,C,C,I,A,T,O,R,E,\$	5	55
#,S,P,E,C,U,L,A,T,O,R,E,\$	5	11
#,S,P,E,R,I,M,E,N,T,A,T,O,R,E,\$	5	7
#,S,Q,U,A,R,T,A,T,O,R,E,\$	5	13
#,S,T,E,R,I,L,I,Z,Z,A,T,O,R,E,\$	5	2
#,S,T,R,O,N,C,A,T,O,R,E,\$	5	50
#,S,T,U,P,R,A,T,O,R,E,\$	5	14
#,S,U,C,C,E,S,S,O,R,E,\$	5	49
#,S,U,G,G,E,R,I,T,O,R,E,\$	5	14
#,S,U,O,N,A,T,O,R,E,\$	5	16
#,T,R,A,D,I,T,O,R,E,\$	5	26
#,T,R,A,D,U,T,T,O,R,E,\$	5	6
#,T,R,A,S,G,R,E,S,S,O,R,E,\$	5	4
#,T,R,U,C,C,A,T,O,R,E,\$	5	3
#,T,R,U,F,F,A,T,O,R,E,\$	5	13
#,U,N,T,O,R,E,\$	5	4
#,U,R,L,A,T,O,R,E,\$	5	2
#,U,T,I,L,I,Z,Z,A,T,O,R,E,\$	5	8
#,V,E,N,D,I,C,A,T,O,R,E,\$	5	3
#,V,E,N,D,I,T,O,R,E,\$	5	56
#,V,E,N,T,I,L,A,T,O,R,E,\$	5	10
#,V,I,A,G,G,I,A,T,O,R,E,\$	5	64
#,V,I,B,R,A,T,O,R,E,\$	5	6
#,V,I,N,C,I,T,O,R,E,\$	5	146
#,V,I,O,L,E,N,T,A,T,O,R,E,\$	5	14
#,V,I,S,I,T,A,T,O,R,E,\$	5	69
#,V,I,S,O,R,E,\$	5	5
#,V,O,G,A,T,O,R,E,\$	5	3
#,D,A,N,Z,A,T,O,R,E,\$	5	6
#,E,S,T,I,M,A,T,O,R,E,\$	5	11
#,I,N,I,Z,I,A,T,O,R,E,\$	5	4
#,P,O,L,A,R,I,Z,Z,A,T,O,R,E,\$	5	2
#,R,E,A,L,I,Z,Z,A,T,O,R,E,\$	5	3
#,R,E,G,O,L,A,T,O,R,E,\$	5	3
#,S,C,O,P,R,I,T,O,R,E,\$	5	11

#,I,M,B,O,N,I,T,O,R,E,\$	5	4
#,A,M,A,T,O,R,E,\$	5	4
#,A,C,C,A,D,E,M,I,C,O,\$	5	44
#,A,C,Q,U,A,T,I,C,O,\$	5	21
#,I,D,R,I,C,O,\$	5	21
#,A,C,R,O,B,A,T,I,C,O,\$	5	8
#,A,L,C,O,L,I,C,O,\$	5	12
#,A,L,F,A,B,E,T,I,C,O,\$	5	6
#,A,L,L,E,G,O,R,I,C,O,\$	5	6
#,A,L,L,E,R,G,I,C,O,\$	5	18
#,A,N,A,G,R,A,F,I,C,O,\$	5	18
#,A,N,A,L,I,T,I,C,O,\$	5	14
#,A,N,A,L,O,G,I,C,O,\$	5	9
#,A,N,A,R,C,H,I,C,O,\$	5	14
#,A,N,A,T,O,M,I,C,O,\$	5	11
#,A,N,E,M,I,C,O,\$	5	5
#,A,N,G,E,L,I,C,O,\$	5	6
#,A,N,T,I,P,A,T,I,C,O,\$	5	20
#,A,N,T,R,O,P,O,L,O,G,I,C,O,\$	5	15
#,A,P,O,C,A,L,I,T,T,I,C,O,\$	5	9
#,A,P,O,S,T,O,L,I,C,O,\$	5	10
#,A,P,P,E,N,N,I,N,I,C,O,\$	5	6
#,P,N,E,U,M,A,T,I,C,O,\$	5	14
#,A,R,I,S,T,O,C,R,A,T,I,C,O,\$	5	39
#,A,R,M,O,N,I,C,O,\$	5	13
#,A,R,O,M,A,T,I,C,O,\$	5	24
#,A,S,B,U,R,G,I,C,O,\$	5	6
#,A,S,C,E,T,I,C,O,\$	5	9
#,A,S,I,A,T,I,C,O,\$	5	25
#,A,S,T,R,O,L,O,G,I,C,O,\$	5	10
#,A,S,T,R,O,N,O,M,I,C,O,\$	5	20
#,A,T,M,O,S,F,E,R,I,C,O,\$	5	45
#,B,E,L,L,I,C,O,\$	5	19
#,A,U,T,O,M,A,T,I,C,O,\$	5	81
#,B,A,L,C,A,N,I,C,O,\$	5	16
#,B,A,L,S,A,M,I,C,O,\$	5	8
#,B,I,B,L,I,C,O,\$	5	30
#,B,I,O,G,R,A,F,I,C,O,\$	5	11
#,B,I,O,L,O,G,I,C,O,\$	5	65
#,B,O,R,B,O,N,I,C,O,\$	5	6
#,B,U,R,O,C,R,A,T,I,C,O,\$	5	56
#,T,E,R,M,I,C,O,\$	5	34
#,C,A,L,O,R,I,C,O,\$	5	11
#,C,A,N,O,N,I,C,O,\$	5	35
#,C,A,O,T,I,C,O,\$	5	22
#,C,A,R,I,S,M,A,T,I,C,O,\$	5	20
#,C,A,R,S,I,C,O,\$	5	7
#,C,A,T,A,S,T,R,O,F,I,C,O,\$	5	23
#,C,A,T,O,D,I,C,O,\$	5	4
#,C,A,T,E,G,O,R,I,C,O,\$	5	31

#,C,E,R,V,E,L,L,O,T,I,C,O,\$	5	7
#,C,H,I,L,O,M,E,T,R,I,C,O,\$	5	7
#,C,H,I,R,U,R,G,I,C,O,\$	5	56
#,C,I,C,L,I,C,O,\$	5	8
#,C,I,L,I,N,D,R,I,C,O,\$	5	7
#,C,L,I,M,A,T,I,C,O,\$	5	19
#,C,I,N,E,M,A,T,O,G,R,A,F,I,C,O,\$	5	154
#,C,R,O,M,A,T,I,C,O,\$	5	17
#,C,O,R,E,O,G,R,A,F,I,C,O,\$	5	7
#,S,O,M,A,T,I,C,O,\$	5	11
#,C,O,S,M,E,T,I,C,O,\$	5	12
#,C,O,S,M,I,C,O,\$	5	26
#,C,R,A,N,I,C,O,\$	5	6
#,C,R,O,N,O,L,O,G,I,C,O,\$	5	10
#,D,E,M,A,G,O,G,I,C,O,\$	5	11
#,D,E,M,O,C,R,A,T,I,C,O,\$	5	290
#,D,I,S,P,O,T,I,C,O,\$	5	4
#,D,I,N,A,S,T,I,C,O,\$	5	4
#,D,E,M,O,S,C,O,P,I,C,O,\$	5	3
#,D,E,M,O,G,R,A,F,I,C,O,\$	5	46
#,D,I,A,B,E,T,I,C,O,\$	5	8
#,D,I,A,G,N,O,S,T,I,C,O,\$	5	10
#,D,I,A,B,O,L,I,C,O,\$	5	36
#,D,I,D,A,S,C,A,L,I,C,O,\$	5	9
#,D,I,E,T,E,T,I,C,O,\$	5	18
#,D,I,P,L,O,M,A,T,I,C,O,\$	5	73
#,D,I,S,C,O,G,R,A,F,I,C,O,\$	5	42
#,D,O,G,M,A,T,I,C,O,\$	5	6
#,D,O,R,I,C,O,\$	5	6
#,D,R,A,M,M,A,T,I,C,O,\$	5	224
#,E,C,O,N,O,M,I,C,O,\$	5	772
#,E,C,O,L,O,G,I,C,O,\$	5	61
#,E,B,R,A,I,C,O,\$	5	104
#,E,L,L,E,N,I,C,O,\$	5	6
#,E,M,B,L,E,M,A,T,I,C,O,\$	5	24
#,E,N,E,R,G,I,C,O,\$	5	31
#,E,N,I,G,M,A,T,I,C,O,\$	5	20
#,E,N,T,U,S,I,A,S,T,I,C,O,\$	5	8
#,E,P,I,D,E,M,I,O,L,O,G,I,C,O,\$	5	8
#,E,P,I,S,O,D,I,C,O,\$	5	7
#,E,R,M,E,T,I,C,O,\$	5	10
#,E,R,O,I,C,O,\$	5	44
#,E,T,N,I,C,O,\$	5	49
#,E,U,F,O,R,I,C,O,\$	5	11
#,F,A,R,A,O,N,I,C,O,\$	5	13
#,F,I,S,I,O,L,O,G,I,C,O,\$	5	31
#,F,R,E,N,E,T,I,C,O,\$	5	31
#,G,A,L,A,T,T,I,C,O,\$	5	6
#,G,A,S,T,R,O,N,O,M,I,C,O,\$	5	20
#,G,E,N,I,C,O,\$	5	8

#,G,E,N,E,R,I,C,O,\$	5	40
#,G,E,O,L,O,G,I,C,O,\$	5	10
#,G,E,O,M,E,T,R,I,C,O,\$	5	23
#,G,E,R,A,R,C,H,I,C,O,\$	5	10
#,T,E,U,T,O,N,I,C,O,\$	5	5
#,N,I,P,P,O,N,I,C,O,\$	5	22
#,G,I,N,E,C,O,L,O,G,I,C,O,\$	5	12
#,L,U,D,I,C,O,\$	5	8
#,G,R,A,N,I,T,I,C,O,\$	5	10
#,I,G,I,E,N,I,C,O,\$	5	38
#,I,N,I,Z,I,A,T,I,C,O,\$	5	10
#,I,O,N,I,C,O,\$	5	7
#,I,P,O,T,E,T,I,C,O,\$	5	27
#,I,S,T,E,R,I,C,O,\$	5	19
#,L,A,T,T,I,C,O,\$	5	5
#,L,E,S,B,I,C,O,\$	5	9
#,L,I,B,I,C,O,\$	5	29
#,L,I,N,F,A,T,I,C,O,\$	5	7
#,L,U,N,A,T,I,C,O,\$	5	5
#,M,A,L,I,N,C,O,N,I,C,O,\$	5	40
#,M,A,S,S,O,N,I,C,O,\$	5	13
#,M,E,D,I,A,N,I,C,O,\$	5	12
#,M,E,L,O,D,I,C,O,\$	5	11
#,M,E,T,A,F,O,R,I,C,O,\$	5	10
#,M,E,T,A,L,L,I,C,O,\$	5	40
#,M,E,T,E,O,R,O,L,O,G,I,C,O,\$	5	28
#,M,E,T,O,D,I,C,O,\$	5	5
#,M,I,C,R,O,S,C,O,P,I,C,O,\$	5	8
#,M,I,T,I,C,O,\$	5	110
#,M,I,T,O,L,O,G,I,C,O,\$	5	10
#,M,O,N,O,G,R,A,F,I,C,O,\$	5	7
#,N,E,V,R,O,T,I,C,O,\$	5	16
#,N,O,R,D,I,C,O,\$	5	22
#,N,O,S,T,A,L,G,I,C,O,\$	5	23
#,N,U,C,L,E,I,C,O,\$	5	16
#,N,U,M,E,R,I,C,O,\$	5	23
#,O,D,O,N,T,O,I,A,T,R,I,C,O,\$	5	4
#,O,L,I,M,P,I,C,O,\$	5	30
#,O,M,E,R,I,C,O,\$	5	6
#,O,R,G,A,N,I,C,O,\$	5	56
#,O,R,T,O,P,E,D,I,C,O,\$	5	4
#,P,A,N,C,R,E,A,T,I,C,O,\$	5	4
#,P,A,N,O,R,A,M,I,C,O,\$	5	16
#,P,A,R,A,N,O,I,C,O,\$	5	4
#,P,A,R,T,I,T,I,C,O,\$	5	13
#,P,A,T,O,L,O,G,I,C,O,\$	5	10
#,P,E,D,A,G,O,G,I,C,O,\$	5	16
#,P,E,R,I,F,E,R,I,C,O,\$	5	36
#,P,E,R,I,O,D,I,C,O,\$	5	24
#,I,T,T,I,C,O,\$	5	20

#,P,L,A,T,O,N,I,C,O,\$	5	12
#,P,O,E,T,I,C,O,\$	5	49
#,P,O,L,I,F,O,N,I,C,O,\$	5	5
#,P,R,O,G,R,A,M,M,A,T,I,C,O,\$	5	29
#,P,R,O,S,A,I,C,O,\$	5	5
#,P,S,I,C,H,I,C,O,\$	5	78
#,P,S,I,C,H,E,D,E,L,I,C,O,\$	5	7
#,P,S,I,C,H,I,A,T,R,I,C,O,\$	5	31
#,P,S,I,C,O,L,O,G,I,C,O,\$	5	134
#,R,I,T,M,I,C,O,\$	5	13
#,S,A,T,I,R,I,C,O,\$	5	20
#,S,C,A,R,A,M,A,N,T,I,C,O,\$	5	7
#,S,C,E,N,I,C,O,\$	5	19
#,S,C,H,E,L,E,T,R,I,C,O,\$	5	4
#,S,C,H,E,M,A,T,I,C,O,\$	5	6
#,S,C,H,I,Z,O,F,R,E,N,I,C,O,\$	5	6
#,S,C,O,R,B,U,T,I,C,O,\$	5	6
#,S,E,L,V,A,T,I,C,O,\$	5	34
#,S,F,E,R,I,C,O,\$	5	5
#,S,I,D,E,R,U,R,G,I,C,O,\$	5	9
#,S,I,M,B,O,L,I,C,O,\$	5	76
#,A,S,I,M,M,E,T,R,I,C,O,\$	5	9
#,S,I,M,P,A,T,I,C,O,\$	5	127
#,S,I,N,T,A,T,T,I,C,O,\$	5	5
#,S,I,N,T,E,T,I,C,O,\$	5	28
#,S,I,N,T,O,M,A,T,I,C,O,\$	5	10
#,S,I,S,M,I,C,O,\$	5	5
#,S,I,S,T,E,M,A,T,I,C,O,\$	5	32
#,S,O,C,I,O,L,O,G,I,C,O,\$	5	15
#,O,N,I,R,I,C,O,\$	5	25
#,I,B,E,R,I,C,O,\$	5	13
#,S,P,A,S,T,I,C,O,\$	5	5
#,P,R,E,I,S,T,O,R,I,C,O,\$	5	17
#,S,T,O,R,I,C,O,\$	5	468
#,S,T,R,A,T,O,S,F,E,R,I,C,O,\$	5	12
#,T,E,L,E,F,O,N,I,C,O,\$	5	123
#,T,E,M,A,T,I,C,O,\$	5	7
#,C,R,O,N,I,C,O,\$	5	28
#,T,E,O,R,I,C,O,\$	5	62
#,T,E,R,A,P,E,U,T,I,C,O,\$	5	46
#,T,I,P,O,G,R,A,F,I,C,O,\$	5	9
#,T,I,T,A,N,I,C,O,\$	5	7
#,T,O,N,I,C,O,\$	5	8
#,V,I,T,A,M,I,N,I,C,O,\$	5	5
#,V,U,L,C,A,N,I,C,O,\$	5	26
#,A,M,L,E,T,I,C,O,\$	5	4
#,A,N,T,I,T,E,T,I,C,O,\$	5	5
#,A,N,T,O,L,O,G,I,C,O,\$	5	4
#,C,O,L,O,N,I,C,O,\$	5	5
#,C,O,L,L,E,R,I,C,O,\$	5	4

#,I,R,O,N,I,C,O,\$
#,E,N,F,A,T,I,C,O,\$

5
5

57
2

tsom input

frequenze bilanciate

#,A,B,I,T,A,\$	5
#,A,B,I,T,A,B,I,L,E,\$	5
#,A,B,I,T,A,N,O,\$	5
#,A,B,I,T,A,N,T,E,\$	5
#,A,B,I,T,A,T,O,\$	5
#,A,B,I,T,A,T,O,R,E,\$	5
#,A,B,I,T,A,N,D,O,\$	5
#,A,B,I,T,A,Z,I,O,N,E,\$	5
#,A,B,I,T,I,A,M,O,\$	5
#,A,B,I,T,O,\$	5
#,A,D,E,R,E,N,D,O,\$	5
#,A,D,E,R,E,N,T,E,\$	5
#,A,D,E,R,E,N,Z,A,\$	5
#,A,D,E,R,I,S,C,E,\$	5
#,A,D,E,R,I,S,C,O,N,O,\$	5
#,A,D,E,R,I,T,O,\$	5
#,A,D,E,R,I,V,A,N,O,\$	5
#,A,D,E,S,I,O,N,E,\$	5
#,A,D,E,S,I,V,O,\$	5
#,A,D,E,R,I,R,V,I,\$	5
#,A,G,G,R,E,D,E,N,D,O,\$	5
#,A,G,G,R,E,D,I,S,C,E,\$	5
#,A,G,G,R,E,D,I,S,C,O,N,O,\$	5
#,A,G,G,R,E,D,I,T,O,\$	5
#,A,G,G,R,E,D,I,V,A,\$	5
#,A,G,G,R,E,S,S,I,O,N,E,\$	5
#,A,G,G,R,E,S,S,I,V,A,M,E,N,T,E,\$	5
#,A,G,G,R,E,D,I,R,E,\$	5
#,A,G,G,R,E,S,S,I,V,O,\$	5
#,A,G,G,R,E,S,S,O,R,E,\$	5
#,A,M,A,\$	5
#,A,M,A,B,I,L,E,\$	5
#,A,M,A,N,O,\$	5
#,A,M,A,N,T,E,\$	5
#,A,M,A,T,E,\$	5
#,A,M,A,T,O,\$	5
#,A,M,A,T,O,R,E,\$	5
#,A,M,A,V,A,\$	5
#,A,M,I,A,M,O,\$	5
#,A,M,O,\$	5
#,A,M,M,I,R,A,\$	5
#,A,M,M,I,R,A,B,I,L,E,\$	5
#,A,M,M,I,R,A,N,O,\$	5
#,A,M,M,I,R,A,T,O,\$	5
#,A,M,M,I,R,A,T,O,R,E,\$	5
#,A,M,M,I,R,A,T,R,I,C,E,\$	5
#,A,M,M,I,R,A,Z,I,O,N,E,\$	5
#,A,M,M,I,R,E,V,O,L,E,\$	5

#,A,M,M,I,R,I,A,M,O,\$	5
#,A,M,M,I,R,O,\$	5
#,A,P,P,L,I,C,A,\$	5
#,A,P,P,L,I,C,A,B,I,L,E,\$	5
#,A,P,P,L,I,C,A,T,O,R,E,\$	5
#,A,P,P,L,I,C,A,N,D,O,\$	5
#,A,P,P,L,I,C,A,N,O,\$	5
#,A,P,P,L,I,C,A,T,I,V,O,\$	5
#,A,P,P,L,I,C,A,T,O,\$	5
#,A,P,P,L,I,C,A,Z,I,O,N,E,\$	5
#,A,P,P,L,I,C,H,I,A,M,O,\$	5
#,A,P,P,L,I,C,O,\$	5
#,C,O,L,L,A,B,O,R,A,\$	5
#,C,O,L,L,A,B,O,R,A,N,D,O,\$	5
#,C,O,L,L,A,B,O,R,A,N,O,\$	5
#,C,O,L,L,A,B,O,R,A,N,T,E,\$	5
#,C,O,L,L,A,B,O,R,A,T,I,V,O,\$	5
#,C,O,L,L,A,B,O,R,A,T,O,\$	5
#,C,O,L,L,A,B,O,R,A,T,O,R,E,\$	5
#,C,O,L,L,A,B,O,R,A,T,R,I,C,E,\$	5
#,C,O,L,L,A,B,O,R,A,V,A,\$	5
#,C,O,L,L,A,B,O,R,A,Z,I,O,N,E,\$	5
#,C,O,M,P,E,T,E,\$	5
#,C,O,M,P,E,T,E,N,T,E,\$	5
#,C,O,M,P,E,T,E,N,Z,A,\$	5
#,C,O,M,P,E,T,I,A,M,O,\$	5
#,C,O,M,P,E,T,E,R,E,\$	5
#,C,O,M,P,E,T,I,T,I,V,O,\$	5
#,C,O,M,P,E,T,I,T,O,R,E,\$	5
#,C,O,M,P,E,T,I,Z,I,O,N,E,\$	5
#,C,O,M,P,E,T,O,N,O,\$	5
#,C,O,M,U,N,I,C,A,\$	5
#,C,O,M,U,N,I,C,A,N,D,O,\$	5
#,C,O,M,U,N,I,C,A,N,O,\$	5
#,C,O,M,U,N,I,C,A,N,T,E,\$	5
#,C,O,M,U,N,I,C,A,T,I,V,O,\$	5
#,C,O,M,U,N,I,C,A,T,O,\$	5
#,C,O,M,U,N,I,C,A,T,O,R,E,\$	5
#,C,O,M,U,N,I,C,A,V,A,\$	5
#,C,O,M,U,N,I,C,A,V,O,\$	5
#,C,O,M,U,N,I,C,A,Z,I,O,N,E,\$	5
#,C,O,N,C,E,T,T,O,\$	5
#,C,O,N,C,E,P,I,A,M,O,\$	5
#,C,O,N,C,E,P,I,B,I,L,E,\$	5
#,C,O,N,C,E,P,I,M,E,N,T,O,\$	5
#,C,O,N,C,E,P,I,S,C,E,\$	5
#,C,O,N,C,E,P,I,S,C,O,N,O,\$	5
#,C,O,N,C,E,P,I,R,E,\$	5
#,C,O,N,C,E,P,I,T,O,\$	5
#,C,O,N,C,E,P,I,V,A,\$	5

#,C,O,N,C,E,Z,I,O,N,E,\$	5
#,C,R,E,A,\$	5
#,C,R,E,A,N,O,\$	5
#,C,R,E,A,N,Z,A,\$	5
#,C,R,E,A,T,I,V,O,\$	5
#,C,R,E,A,T,O,\$	5
#,C,R,E,A,T,O,R,E,\$	5
#,C,R,E,A,T,R,I,C,E,\$	5
#,C,R,E,A,T,U,R,A,\$	5
#,C,R,E,A,Z,I,O,N,E,\$	5
#,C,R,E,O,\$	5
#,C,R,E,I,A,M,O,\$	5
#,S,E,P,A,R,A,Z,I,O,N,E,\$	5
#,S,E,P,A,R,A,T,A,M,E,N,T,E,\$	5
#,S,E,P,A,R,A,T,O,R,E,\$	5
#,S,E,P,A,R,A,N,O,\$	5
#,S,E,P,A,R,A,T,O,\$	5
#,S,E,P,A,R,A,R,E,\$	5
#,S,E,P,A,R,A,V,A,\$	5
#,S,E,P,A,R,A,T,E,Z,Z,A,\$	5
#,S,E,P,A,R,A,R,O,N,O,\$	5
#,S,E,P,A,R,A,T,E,Z,Z,A,\$	5
#,S,E,P,A,R,A,T,I,S,M,O,\$	5
#,S,E,P,A,R,A,R,I,A,M,O,\$	5
#,D,E,T,E,R,M,I,N,A,\$	5
#,D,E,T,E,R,M,I,N,A,N,D,O,\$	5
#,D,E,T,E,R,M,I,N,A,N,O,\$	5
#,D,E,T,E,R,M,I,N,A,N,T,E,\$	5
#,D,E,T,E,R,M,I,N,A,T,O,\$	5
#,D,E,T,E,R,M,I,N,A,Z,I,O,N,E,\$	5
#,D,E,T,E,R,M,I,N,A,R,E,\$	5
#,D,E,T,E,R,M,I,N,I,\$	5
#,D,E,T,E,R,M,I,N,I,S,M,O,\$	5
#,D,E,T,E,R,M,I,N,A,T,O,R,E,\$	5
#,D,I,M,O,S,T,R,A,\$	5
#,D,I,M,O,S,T,R,A,T,O,R,E,\$	5
#,D,I,M,O,S,T,R,A,B,I,L,E,\$	5
#,D,I,M,O,S,T,R,A,N,D,O,\$	5
#,D,I,M,O,S,T,R,A,N,O,\$	5
#,D,I,M,O,S,T,R,A,N,T,E,\$	5
#,D,I,M,O,S,T,R,A,T,I,V,O,\$	5
#,D,I,M,O,S,T,R,A,T,O,\$	5
#,D,I,M,O,S,T,R,A,V,A,\$	5
#,D,I,M,O,S,T,R,A,Z,I,O,N,E,\$	5
#,D,I,M,O,S,T,R,I,\$	5
#,D,I,S,P,O,N,E,\$	5
#,D,I,S,P,O,N,G,O,\$	5
#,D,I,S,P,O,N,G,O,N,O,\$	5
#,D,I,S,P,O,N,I,A,M,O,\$	5
#,D,I,S,P,O,N,I,B,I,L,E,\$	5

#,D,I,S,P,O,N,E,T,E,\$	5
#,D,I,S,P,O,R,R,E,\$	5
#,D,I,S,P,O,S,I,T,I,V,O,\$	5
#,D,I,S,P,O,S,I,Z,I,O,N,E,\$	5
#,D,I,S,P,O,S,T,O,\$	5
#,E,C,C,I,T,A,B,I,L,E,\$	5
#,E,C,C,I,T,A,R,E,\$	5
#,E,C,C,I,T,A,M,E,N,T,O,\$	5
#,E,C,C,I,T,A,N,O,\$	5
#,E,C,C,I,T,A,N,T,E,\$	5
#,E,C,C,I,T,A,T,O,\$	5
#,E,C,C,I,T,A,V,A,\$	5
#,E,C,C,I,T,A,V,A,N,O,\$	5
#,E,C,C,I,T,A,Z,I,O,N,E,\$	5
#,E,D,U,C,A,N,D,A,\$	5
#,E,D,U,C,A,N,D,O,\$	5
#,E,D,U,C,A,T,A,\$	5
#,E,D,U,C,A,T,E,\$	5
#,E,D,U,C,H,I,A,M,O,\$	5
#,E,D,U,C,A,R,E,\$	5
#,E,D,U,C,A,T,I,V,O,\$	5
#,E,D,U,C,A,T,O,\$	5
#,E,D,U,C,A,T,O,R,E,\$	5
#,E,D,U,C,A,T,R,I,C,E,\$	5
#,E,D,U,C,A,Z,I,O,N,E,\$	5
#,C,O,N,T,E,S,T,A,Z,I,O,N,E,\$	5
#,C,O,N,T,E,S,T,A,T,O,R,E,\$	5
#,C,O,N,T,E,S,T,A,B,I,L,E,\$	5
#,C,O,N,T,E,S,T,A,T,O,\$	5
#,C,O,N,T,E,S,T,A,N,O,\$	5
#,C,O,N,T,E,S,T,A,N,D,O,\$	5
#,C,O,N,T,E,S,T,A,V,A,N,O,\$	5
#,C,O,N,T,E,S,T,A,\$	5
#,C,O,N,T,E,S,T,O,\$	5
#,F,O,N,D,A,\$	5
#,F,O,N,D,A,M,E,N,T,A,L,E,\$	5
#,F,O,N,D,A,M,E,N,T,A,L,I,S,M,O,\$	5
#,F,O,N,D,A,M,E,N,T,A,L,I,S,T,A,\$	5
#,F,O,N,D,A,M,E,N,T,O,\$	5
#,F,O,N,D,A,R,E,\$	5
#,F,O,N,D,A,N,O,\$	5
#,F,O,N,D,A,T,E,Z,Z,A,\$	5
#,F,O,N,D,A,T,O,\$	5
#,F,O,N,D,A,T,O,R,E,\$	5
#,F,O,N,D,A,T,R,I,C,E,\$	5
#,F,O,N,D,A,Z,I,O,N,E,\$	5
#,F,O,R,M,A,L,E,\$	5
#,F,O,R,M,A,R,E,\$	5
#,F,O,R,M,A,N,D,O,\$	5
#,F,O,R,M,A,N,O,\$	5

#,F,O,R,M,A,T,I,V,O,\$	5
#,F,O,R,M,A,T,O,\$	5
#,F,O,R,M,A,Z,I,O,N,E,\$	5
#,F,O,R,M,A,T,O,R,E,\$	5
#,F,O,R,M,I,A,M,O,\$	5
#,F,O,R,M,O,S,O,\$	5
#,I,L,L,U,D,E,\$	5
#,I,L,L,U,D,E,V,O,\$	5
#,I,L,L,U,D,E,R,E,\$	5
#,I,L,L,U,D,I,A,M,O,\$	5
#,I,L,L,U,D,O,N,O,\$	5
#,I,L,L,U,S,I,O,N,E,\$	5
#,I,L,L,U,S,I,O,N,I,S,M,O,\$	5
#,I,L,L,U,S,I,O,N,I,S,T,A,\$	5
#,I,L,L,U,S,I,O,N,I,S,T,I,C,O,\$	5
#,I,L,L,U,S,O,\$	5
#,I,L,L,U,S,O,R,I,O,\$	5
#,I,M,M,A,G,I,N,A,\$	5
#,I,M,M,A,G,I,N,I,A,M,O,\$	5
#,I,M,M,A,G,I,N,A,B,I,L,E,\$	5
#,I,M,M,A,G,I,N,A,R,I,O,\$	5
#,I,M,M,A,G,I,N,A,T,O,\$	5
#,I,M,M,A,G,I,N,A,V,A,\$	5
#,I,M,M,A,G,I,N,A,V,A,N,O,\$	5
#,I,M,M,A,G,I,N,A,V,O,\$	5
#,I,M,M,A,G,I,N,A,Z,I,O,N,E,\$	5
#,I,M,M,A,G,I,N,I,A,M,O,\$	5
#,I,M,M,A,G,I,N,O,\$	5
#,I,N,D,I,C,A,\$	5
#,I,N,D,I,C,H,I,A,M,O,\$	5
#,I,N,D,I,C,A,T,E,\$	5
#,I,N,D,I,C,A,N,D,O,\$	5
#,I,N,D,I,C,A,N,O,\$	5
#,I,N,D,I,C,A,T,I,V,O,\$	5
#,I,N,D,I,C,A,T,O,\$	5
#,I,N,D,I,C,A,T,O,R,E,\$	5
#,I,N,D,I,C,A,V,A,\$	5
#,I,N,D,I,C,A,Z,I,O,N,E,\$	5
#,I,N,D,I,C,H,I,\$	5
#,I,N,D,I,C,O,\$	5
#,I,N,F,O,R,M,A,\$	5
#,I,N,F,O,R,M,A,R,E,\$	5
#,I,N,F,O,R,M,I,A,M,O,\$	5
#,I,N,F,O,R,M,A,T,E,\$	5
#,I,N,F,O,R,M,A,N,O,\$	5
#,I,N,F,O,R,M,A,T,I,C,O,\$	5
#,I,N,F,O,R,M,A,T,I,V,O,\$	5
#,I,N,F,O,R,M,A,T,O,\$	5
#,I,N,F,O,R,M,A,T,O,R,E,\$	5
#,I,N,F,O,R,M,A,Z,I,O,N,E,\$	5

#,I,N,F,O,R,M,O,\$	5
#,I,N,I,Z,I,A,\$	5
#,I,N,I,Z,I,A,M,O,\$	5
#,I,N,I,Z,I,A,T,E,\$	5
#,I,N,I,Z,I,A,L,E,\$	5
#,I,N,I,Z,I,A,L,M,E,N,T,E,\$	5
#,I,N,I,Z,I,A,M,O,\$	5
#,I,N,I,Z,I,A,N,O,\$	5
#,I,N,I,Z,I,A,T,I,C,O,\$	5
#,I,N,I,Z,I,A,T,I,V,A,\$	5
#,I,N,I,Z,I,A,T,O,\$	5
#,I,N,I,Z,I,A,T,O,R,E,\$	5
#,I,N,I,Z,I,A,V,A,N,O,\$	5
#,I,N,I,Z,I,A,Z,I,O,N,E,\$	5
#,I,N,I,Z,I,O,\$	5
#,I,N,T,E,G,R,A,R,E,\$	5
#,I,N,T,E,G,R,A,T,E,\$	5
#,I,N,T,E,G,R,I,A,M,O,\$	5
#,I,N,T,E,G,R,A,L,E,\$	5
#,I,N,T,E,G,R,A,L,I,S,M,O,\$	5
#,I,N,T,E,G,R,A,L,I,S,T,A,\$	5
#,I,N,T,E,G,R,A,L,M,E,N,T,E,\$	5
#,I,N,T,E,G,R,A,N,O,\$	5
#,I,N,T,E,G,R,A,N,T,E,\$	5
#,I,N,T,E,G,R,A,T,O,R,E,\$	5
#,I,N,T,E,G,R,A,Z,I,O,N,E,\$	5
#,I,N,T,E,G,R,O,\$	5
#,I,N,T,E,R,P,R,E,T,A,\$	5
#,I,N,T,E,R,P,R,E,T,I,A,M,O,\$	5
#,I,N,T,E,R,P,R,E,T,A,B,I,L,E,\$	5
#,I,N,T,E,R,P,R,E,T,A,N,D,O,\$	5
#,I,N,T,E,R,P,R,E,T,A,N,O,\$	5
#,I,N,T,E,R,P,R,E,T,A,T,I,V,O,\$	5
#,I,N,T,E,R,P,R,E,T,A,T,O,\$	5
#,I,N,T,E,R,P,R,E,T,A,V,A,\$	5
#,I,N,T,E,R,P,R,E,T,A,Z,I,O,N,E,\$	5
#,I,N,T,E,R,P,R,E,T,E,\$	5
#,I,N,T,E,R,P,R,E,T,O,\$	5
#,I,N,T,E,R,R,O,G,A,\$	5
#,I,N,T,E,R,R,O,G,A,N,D,O,\$	5
#,I,N,T,E,R,R,O,G,A,N,O,\$	5
#,I,N,T,E,R,R,O,G,A,T,I,V,O,\$	5
#,I,N,T,E,R,R,O,G,A,T,O,\$	5
#,I,N,T,E,R,R,O,G,A,T,O,R,E,\$	5
#,I,N,T,E,R,R,O,G,A,T,O,R,I,O,\$	5
#,I,N,T,E,R,R,O,G,A,V,A,\$	5
#,I,N,T,E,R,R,O,G,A,T,E,\$	5
#,I,N,T,E,R,R,O,G,A,Z,I,O,N,E,\$	5
#,I,N,T,E,R,R,O,G,O,\$	5
#,I,N,T,U,E,N,D,O,\$	5

#,I,N,T,U,I,B,I,L,E,\$	5
#,I,N,T,U,I,S,C,E,\$	5
#,I,N,T,U,I,S,C,O,\$	5
#,I,N,T,U,I,S,C,O,N,O,\$	5
#,I,N,T,U,I,T,I,V,O,\$	5
#,I,N,T,U,I,T,O,\$	5
#,I,N,T,U,I,V,A,\$	5
#,I,N,T,U,I,V,O,\$	5
#,I,N,T,U,I,Z,I,O,N,E,\$	5
#,I,R,R,I,T,A,B,I,L,E,\$	5
#,I,R,R,I,T,A,R,E,\$	5
#,I,R,R,I,T,I,A,M,O,\$	5
#,I,R,R,I,T,A,N,T,E,\$	5
#,I,R,R,I,T,A,R,L,O,\$	5
#,I,R,R,I,T,A,R,M,I,\$	5
#,I,R,R,I,T,A,T,O,\$	5
#,I,R,R,I,T,A,Z,I,O,N,E,\$	5
#,L,A,V,O,R,A,\$	5
#,L,A,V,O,R,A,T,E,\$	5
#,L,A,V,O,R,A,N,O,\$	5
#,L,A,V,O,R,A,N,T,E,\$	5
#,L,A,V,O,R,A,T,I,V,O,\$	5
#,L,A,V,O,R,A,T,O,\$	5
#,L,A,V,O,R,A,T,O,R,E,\$	5
#,L,A,V,O,R,A,T,R,I,C,E,\$	5
#,L,A,V,O,R,A,V,A,N,O,\$	5
#,L,A,V,O,R,A,V,O,\$	5
#,L,A,V,O,R,A,Z,I,O,N,E,\$	5
#,L,A,V,O,R,I,A,M,O,\$	5
#,L,A,V,O,R,O,\$	5
#,L,E,G,G,E,\$	5
#,L,E,G,G,E,R,E,\$	5
#,L,E,G,G,E,N,D,O,\$	5
#,L,E,G,G,I,A,M,O,\$	5
#,L,E,G,G,I,O,\$	5
#,L,E,G,G,O,\$	5
#,L,E,G,G,O,N,O,\$	5
#,L,E,T,T,O,\$	5
#,L,E,T,T,O,R,E,\$	5
#,L,E,T,T,R,I,C,E,\$	5
#,L,E,T,T,U,R,A,\$	5
#,L,I,M,I,T,A,\$	5
#,L,I,M,I,T,A,R,E,\$	5
#,L,I,M,I,T,A,I,\$	5
#,L,I,M,I,T,A,N,D,O,\$	5
#,L,I,M,I,T,A,R,S,I,\$	5
#,L,I,M,I,T,A,T,A,M,E,N,T,E,\$	5
#,L,I,M,I,T,A,T,I,V,O,\$	5
#,L,I,M,I,T,A,T,O,\$	5
#,L,I,M,I,T,A,V,A,\$	5

#,L,I,M,I,T,A,Z,I,O,N,E,\$	5
#,L,I,M,I,T,O,\$	5
#,A,M,M,I,N,I,S,T,R,A,\$	5
#,A,M,M,I,N,I,S,T,R,A,T,I,V,O,\$	5
#,A,M,M,I,N,I,S,T,R,A,T,O,\$	5
#,A,M,M,I,N,I,S,T,R,A,T,O,R,E,\$	5
#,A,M,M,I,N,I,S,T,R,A,R,E,\$	5
#,A,M,M,I,N,I,S,T,R,A,Z,I,O,N,E,\$	5
#,M,I,N,I,S,T,E,R,I,A,L,E,\$	5
#,M,I,N,I,S,T,E,R,O,\$	5
#,M,O,B,I,L,E,\$	5
#,M,U,O,V,E,R,E,\$	5
#,M,O,B,I,L,I,T,A,Z,I,O,N,E,\$	5
#,M,O,S,S,O,\$	5
#,M,O,T,O,R,E,\$	5
#,M,O,T,O,R,I,O,\$	5
#,M,O,V,I,M,E,N,T,O,\$	5
#,M,U,O,V,E,\$	5
#,M,U,O,V,I,A,M,O,\$	5
#,M,U,O,V,O,\$	5
#,M,U,O,V,O,N,O,\$	5
#,M,U,T,A,N,T,E,\$	5
#,M,U,T,A,M,E,N,T,O,\$	5
#,M,U,T,A,N,O,\$	5
#,M,U,T,A,N,O,\$	5
#,M,U,T,A,R,E,\$	5
#,M,U,T,A,T,O,\$	5
#,M,U,T,A,R,E,\$	5
#,M,U,T,A,Z,I,O,N,E,\$	5
#,M,U,T,E,V,O,L,E,\$	5
#,N,U,T,R,E,\$	5
#,N,U,T,R,I,E,N,T,E,\$	5
#,N,U,T,R,I,M,E,N,T,O,\$	5
#,N,U,T,R,I,A,M,O,\$	5
#,N,U,T,R,I,T,I,V,O,\$	5
#,N,U,T,R,I,T,O,\$	5
#,N,U,T,R,I,V,A,\$	5
#,N,U,T,R,I,R,E,\$	5
#,N,U,T,R,I,Z,I,O,N,E,\$	5
#,N,U,T,R,O,N,O,\$	5
#,O,P,E,R,A,\$	5
#,O,P,E,R,A,I,O,\$	5
#,O,P,E,R,A,N,O,\$	5
#,O,P,E,R,A,N,T,E,\$	5
#,O,P,E,R,A,T,I,V,A,M,E,N,T,E,\$	5
#,O,P,E,R,A,R,E,\$	5
#,O,P,E,R,A,T,O,\$	5
#,O,P,E,R,A,T,O,R,E,\$	5
#,O,P,E,R,A,T,O,R,I,O,\$	5
#,O,P,E,R,A,Z,I,O,N,E,\$	5

#,O,P,E,R,E,T,T,A,\$	5
#,O,P,E,R,O,S,O,\$	5
#,O,S,S,E,R,V,A,\$	5
#,O,S,S,E,R,V,A,R,E,\$	5
#,O,S,S,E,R,V,A,N,O,\$	5
#,O,S,S,E,R,V,A,N,T,E,\$	5
#,O,S,S,E,R,V,A,N,Z,A,\$	5
#,O,S,S,E,R,V,A,T,I,V,O,\$	5
#,O,S,S,E,R,V,A,T,O,R,E,\$	5
#,O,S,S,E,R,V,A,T,O,R,I,O,\$	5
#,O,S,S,E,R,V,A,Z,I,O,N,E,\$	5
#,O,S,S,E,R,V,I,A,M,O,\$	5
#,P,R,E,P,A,R,A,\$	5
#,P,R,E,P,A,R,A,N,D,O,\$	5
#,P,R,E,P,A,R,A,R,E,\$	5
#,P,R,E,P,A,R,A,T,O,\$	5
#,P,R,E,P,A,R,A,T,O,R,E,\$	5
#,P,R,E,P,A,R,A,T,O,R,I,O,\$	5
#,P,R,E,P,A,R,A,V,A,\$	5
#,P,R,E,P,A,R,A,Z,I,O,N,E,\$	5
#,P,R,E,P,A,R,O,\$	5
#,P,R,E,S,E,N,T,A,\$	5
#,P,R,E,S,E,N,T,A,B,I,L,E,\$	5
#,P,R,E,S,E,N,T,A,N,D,O,\$	5
#,P,R,E,S,E,N,T,A,N,O,\$	5
#,P,R,E,S,E,N,T,A,T,O,\$	5
#,P,R,E,S,E,N,T,A,T,O,R,E,\$	5
#,P,R,E,S,E,N,T,A,R,E,\$	5
#,P,R,E,S,E,N,T,A,Z,I,O,N,E,\$	5
#,P,R,E,S,E,N,T,I,A,M,O,\$	5
#,P,R,E,S,E,N,T,O,\$	5
#,P,R,O,D,O,T,T,O,\$	5
#,P,R,O,D,U,C,E,\$	5
#,P,R,O,D,U,C,E,V,A,\$	5
#,P,R,O,D,U,C,I,A,M,O,\$	5
#,P,R,O,D,U,C,O,N,O,\$	5
#,P,R,O,D,U,R,R,E,\$	5
#,P,R,O,D,U,T,T,I,V,O,\$	5
#,P,R,O,D,U,T,T,O,R,E,\$	5
#,P,R,O,D,U,C,O,N,O,\$	5
#,P,R,O,D,U,Z,I,O,N,E,\$	5
#,P,R,O,G,E,T,T,A,\$	5
#,P,R,O,G,E,T,T,A,N,D,O,\$	5
#,P,R,O,G,E,T,T,A,N,O,\$	5
#,P,R,O,G,E,T,T,A,T,O,\$	5
#,P,R,O,G,E,T,T,A,V,A,\$	5
#,P,R,O,G,E,T,T,A,Z,I,O,N,E,\$	5
#,P,R,O,G,E,T,T,I,A,M,O,\$	5
#,P,R,O,G,E,T,T,O,\$	5
#,P,R,O,G,E,T,T,A,R,E,\$	5

#,P,R,O,G,R,A,M,M,A,B,I,L,E,\$	5
#,P,R,O,G,R,A,M,M,A,N,D,O,\$	5
#,P,R,O,G,R,A,M,M,A,T,E,\$	5
#,P,R,O,G,R,A,M,M,A,T,I,C,O,\$	5
#,P,R,O,G,R,A,M,M,A,T,O,\$	5
#,P,R,O,G,R,A,M,M,A,T,O,R,E,\$	5
#,P,R,O,G,R,A,M,M,A,R,E,\$	5
#,P,R,O,G,R,A,M,M,A,R,E,\$	5
#,P,R,O,G,R,A,M,M,A,Z,I,O,N,E,\$	5
#,P,R,O,G,R,A,M,M,O,\$	5
#,P,R,O,T,E,G,G,E,\$	5
#,P,R,O,T,E,G,G,E,R,E,\$	5
#,P,R,O,T,E,G,G,E,V,A,\$	5
#,P,R,O,T,E,G,G,O,\$	5
#,P,R,O,T,E,G,G,O,N,O,\$	5
#,P,R,O,T,E,T,T,I,V,O,\$	5
#,P,R,O,T,E,T,T,O,\$	5
#,P,R,O,T,E,T,T,O,R,E,\$	5
#,P,R,O,T,E,Z,I,O,N,E,\$	5
#,P,R,O,T,E,G,G,I,A,M,O,\$	5
#,P,R,O,T,E,G,G,E,T,E,\$	5
#,P,R,O,V,O,C,A,\$	5
#,P,R,O,V,O,C,A,R,E,\$	5
#,P,R,O,V,O,C,A,N,D,O,\$	5
#,P,R,O,V,O,C,A,N,O,\$	5
#,P,R,O,V,O,C,A,N,T,E,\$	5
#,P,R,O,V,O,C,A,T,E,\$	5
#,P,R,O,V,O,C,A,T,O,\$	5
#,P,R,O,V,O,C,A,T,O,R,E,\$	5
#,P,R,O,V,O,C,H,I,A,M,O,\$	5
#,P,R,O,V,O,C,A,T,O,R,I,O,\$	5
#,P,R,O,V,O,C,A,Z,I,O,N,E,\$	5
#,P,U,N,E,N,D,O,\$	5
#,P,U,N,I,B,I,L,E,\$	5
#,P,U,N,I,R,E,\$	5
#,P,U,N,I,R,L,O,\$	5
#,P,U,N,I,A,M,O,\$	5
#,P,U,N,I,S,C,E,\$	5
#,P,U,N,I,T,I,V,O,\$	5
#,P,U,N,I,T,O,\$	5
#,P,U,N,I,V,A,\$	5
#,P,U,N,I,Z,I,O,N,E,\$	5
#,R,E,A,G,E,N,D,O,\$	5
#,R,E,A,G,I,S,C,A,\$	5
#,R,E,A,G,I,S,C,E,\$	5
#,R,E,A,G,I,S,C,O,N,O,\$	5
#,R,E,A,G,I,T,O,\$	5
#,R,E,A,G,I,S,C,I,\$	5
#,R,E,A,T,T,I,V,O,\$	5
#,R,E,A,T,T,O,R,E,\$	5

#,R,E,A,G,I,R,E,\$	5
#,R,E,A,G,I,A,M,O,\$	5
#,R,E,A,G,I,T,E,\$	5
#,R,E,A,Z,I,O,N,E,\$	5
#,R,I,L,E,V,A,\$	5
#,R,I,L,E,V,A,R,E,\$	5
#,R,I,L,E,V,I,A,M,O,\$	5
#,R,I,L,E,V,A,B,I,L,E,\$	5
#,R,I,L,E,V,A,M,E,N,T,O,\$	5
#,R,I,L,E,V,A,N,O,\$	5
#,R,I,L,E,V,A,N,T,E,\$	5
#,R,I,L,E,V,A,N,Z,A,\$	5
#,R,I,L,E,V,A,T,E,\$	5
#,R,I,L,E,V,A,T,O,R,E,\$	5
#,R,I,L,E,V,A,Z,I,O,N,E,\$	5
#,R,I,L,I,E,V,O,\$	5
#,S,C,E,N,A,R,I,O,\$	5
#,S,C,E,N,A,T,A,\$	5
#,S,C,E,N,E,G,G,I,A,T,O,R,E,\$	5
#,S,C,E,N,A,\$	5
#,S,C,E,N,E,\$	5
#,S,C,E,N,I,C,A,M,E,N,T,E,\$	5
#,S,C,E,N,I,C,O,\$	5
#,S,C,E,N,O,G,R,A,F,I,A,\$	5
#,S,C,E,N,O,G,R,A,F,O,\$	5
#,S,C,E,N,E,G,G,I,A,T,A,\$	5
#,S,U,C,C,E,D,E,\$	5
#,S,U,C,C,E,D,E,R,E,\$	5
#,S,U,C,C,E,D,E,N,D,O,\$	5
#,S,U,C,C,E,D,E,V,A,\$	5
#,S,U,C,C,E,D,O,N,O,\$	5
#,S,U,C,C,E,D,U,T,O,\$	5
#,S,U,C,C,E,S,S,I,O,N,E,\$	5
#,S,U,C,C,E,D,I,A,M,O,\$	5
#,S,U,C,C,E,S,S,I,V,O,\$	5
#,S,U,C,C,E,S,S,O,\$	5
#,S,U,C,C,E,S,S,O,R,E,\$	5
#,T,E,N,T,A,\$	5
#,T,E,N,T,A,N,D,O,\$	5
#,T,E,N,T,A,N,O,\$	5
#,T,E,N,T,A,T,I,V,O,\$	5
#,T,E,N,T,A,T,O,\$	5
#,T,E,N,T,A,T,O,R,E,\$	5
#,T,E,N,T,A,R,E,\$	5
#,T,E,N,T,A,Z,I,O,N,E,\$	5
#,T,E,N,T,I,A,M,O,\$	5
#,T,E,N,T,O,\$	5
#,T,R,A,D,E,N,D,O,\$	5
#,T,R,A,D,I,M,E,N,T,O,\$	5
#,T,R,A,D,I,S,C,E,\$	5

#,T,R,A,D,I,S,C,O,N,O,\$	5
#,T,R,A,D,I,T,O,\$	5
#,T,R,A,D,I,T,O,R,E,\$	5
#,T,R,A,D,I,V,A,\$	5
#,T,R,A,D,I,R,E,\$	5
#,T,R,A,D,I,A,M,O,\$	5
#,T,R,A,D,I,T,E,\$	5
#,T,R,A,D,O,T,T,O,\$	5
#,T,R,A,D,U,C,E,\$	5
#,T,R,A,D,U,C,E,N,D,O,\$	5
#,T,R,A,D,U,C,E,V,A,\$	5
#,T,R,A,D,U,R,R,E,\$	5
#,T,R,A,D,U,C,O,N,O,\$	5
#,T,R,A,D,U,T,T,O,R,E,\$	5
#,T,R,A,D,U,R,R,E,\$	5
#,T,R,A,D,U,Z,I,O,N,E,\$	5
#,V,A,L,U,T,A,T,E,\$	5
#,V,A,L,U,T,A,R,E,\$	5
#,V,A,L,U,T,I,A,M,O,\$	5
#,V,A,L,U,T,A,N,D,O,\$	5
#,V,A,L,U,T,A,N,O,\$	5
#,V,A,L,U,T,A,R,I,O,\$	5
#,V,A,L,U,T,A,T,I,V,O,\$	5
#,V,A,L,U,T,A,T,O,\$	5
#,V,A,L,U,T,A,Z,I,O,N,E,\$	5
#,V,A,L,U,T,E,R,E,M,O,\$	5
#,V,A,L,U,T,I,\$	5
#,V,A,R,I,A,\$	5
#,V,A,R,I,A,B,I,L,E,\$	5
#,V,A,R,I,A,R,E,\$	5
#,V,A,R,I,A,N,O,\$	5
#,V,A,R,I,A,N,T,E,\$	5
#,V,A,R,I,A,T,O,R,E,\$	5
#,V,A,R,I,A,Z,I,O,N,E,\$	5
#,V,A,R,I,A,T,O,\$	5
#,V,A,R,I,O,\$	5
#,V,E,D,E,\$	5
#,V,E,D,I,A,M,O,\$	5
#,V,E,D,O,\$	5
#,V,E,D,O,N,O,\$	5
#,V,E,D,U,T,A,\$	5
#,V,I,S,I,B,I,L,E,\$	5
#,V,E,D,E,R,E,\$	5
#,V,E,D,E,N,T,E,\$	5
#,V,I,S,I,V,O,\$	5
#,V,I,S,O,N,A,R,I,O,\$	5
#,V,I,S,O,R,E,\$	5
#,V,I,S,T,O,\$	5
#,V,E,D,E,T,E,\$	5
#,V,E,D,U,T,O,\$	5

#,V,I,S,I,O,N,E,\$

5

#,V,I,S,U,A,L,E,\$

5


```
#LIBRERIE
```

```
library(lattice)  
library(ggplot2)  
library(ggsignif)  
library(lme4)  
library(lmerTest)  
library(gamm4)  
library(mgcv)  
library(plyr)  
library(dplyr)  
library(itsadug)  
library(car)  
library(cowplot)  
library(RColorBrewer)  
library(plotly)  
library(MASS)  
library(MuMIn)  
library(effects)  
library(effectsize)  
library(sjmisc)  
library(visreg)  
library(GGally)  
library(sjPlot)  
library(imputeTS)  
library(ggpubr)
```

```
# ESPERIMENTO 1: ADDESTRAMENTO A FREQUENZE BILANCIATE #
```

```
#dataframe epoca 100  
load("DER_100.txt")  
dat1<-read.delim("DER_100.txt", sep="\t", quote="", dec=".", stringsAsFactors = TRUE)  
dim(dat1)  
str(dat1)  
summary(dat1)
```

```
#lunghezza delle forme  
dat1$lunghezza <-rowSums(dat1[, c("prefix", "stem", "suffix")])
```

```
#informazioni sulle serie  
load("Serie.txt")  
dat2<-read.delim("Serie.txt", sep="\t", quote="", dec=".", stringsAsFactors = TRUE)  
dat100 <- merge (dat1, dat2, by=c("word"))
```

```
#salvataggio df aggregato  
write.table(dat100, file="dat100.txt", sep="\t", row.names=FALSE, quote=FALSE)
```

```
#colonna aggiuntiva per epoca  
epoca <- (100)  
dat100 <- cbind (dat100, epoca = epoca)  
write.table(dat100, file="dat100.txt", sep="\t", row.names=FALSE, quote=FALSE)  
datEpoca <- rbind (dat100)  
write.table(datEpoca, file="datEpoca.txt", sep="\t", row.names=FALSE, quote=FALSE)  
datEpoca$Epoca<- as.factor(datEpoca$epoca)
```

```
#da ripetere per tutte le epoche da esaminare
```

```

#plot di regressione
ggplot(datEpoca, aes(dist.to.MB, anticip.BMU, group=conj))+ geom_smooth(method= "loess", se = FALSE,
aes(linetype=conj, color=conj), size=1.2)+ coord_cartesian(ylim=c(0, 3))+
ylab("Predizione")+ xlab("distanza da confine di morfema")+ theme_bw()+
theme(text = element_text (size=20, face="bold"), legend.title = element_blank()) +
ggtitle("Epoca finale")+ theme(plot.title = element_text (lineheight=.8, face="bold",
hjust = 0.5, vjust= 0.5), legend.title = element_blank(),
legend.text=element_text(size=16), axis.text=element_text(size=14))

```

```

#epoche intermedie 8-14
load("dat814.txt")
datIntermedie<-read.delim("dat814.txt", sep="\t", quote="", dec=".", stringsAsFactors = TRUE)
datIntermedie$Epoca<- as.factor(datIntermedie$epoca)

```

```

ggplot(datIntermedie, aes(dist.to.MB, anticip.BMU, group=Epoca))+ geom_smooth(method= "loess", se = FALSE,
aes(linetype=Epoca, color=Epoca), size=1.2)+ coord_cartesian(ylim=c(0, 3))+
ylab("Predizione")+ xlab("distanza da confine di morfema")+ theme_bw()+ facet_wrap (~ conj) +
theme(text = element_text (size=20, face="bold"), legend.title = element_blank()) +
ggtitle("epoche: 8-14")+ theme(plot.title = element_text (lineheight=0.8, face="bold",
hjust = 0.5, vjust= 0.5), legend.title = element_blank(),
legend.text=element_text(size=16), axis.text=element_text(size=14))

```

```

#modello non lineare: funzione gam (generalized additive model)
m1 <- gam (anticip.BMU ~ dist.to.MB * conj + s(repetition.id, bs = "re"), data=dat100)
summary(m2)

```

ESPERIMENTO 2: ADDESTRAMENTO CORPUS-BASED

```

load("DER_F_100.txt")
dat1<-read.delim("DER_F_100.txt", sep="\t", quote="", dec=".", stringsAsFactors = TRUE)

```

```

dat1$lunghezza <-rowSums(dat1[, c("prefix", "stem", "suffix")])
epoca <- (100)
dat1 <- cbind (dat1, epoca = epoca)
dat1$Epoca<- as.factor(dat1$epoca)
library(dplyr)
dat1 <- filter (dat1, dat1$symbol != "#")

```

```

Dat100 <- merge (dat1, dat2, by=c("word"))
write.table(Dat100, file="Dat100.txt", sep="\t", row.names=FALSE, quote=FALSE)

```

```

ggplot(Dat100, aes(dist.to.MB, anticip.BMU, group=conj))+ geom_smooth(method= "loess", se = FALSE,
aes(linetype=conj, color=conj), size=1.2)+ coord_cartesian(ylim=c(0, 3))+
ylab("Predizione")+ xlab("distanza da confine di morfema")+ theme_bw()+
facet_wrap (~ Epoca) +
theme(text = element_text (size=20, face="bold"), legend.title = element_blank()) +
ggtitle("epoca finale")+ theme(plot.title = element_text (lineheight=0.8, face="bold",
hjust = 0.5, vjust= 0.5), legend.title = element_blank(),
legend.text=element_text(size=16), axis.text=element_text(size=14))

```

```

m2 <- gam (anticip.BMU ~ dist.to.MB * conj + s(repetition.id, bs = "re"), data=Dat100)
summary(m2)

```

#Considero le allomorfie di -ZIONE

```

DatAllomorph<-read.delim("Dat100.txt", sep="\t", quote="", dec=".", stringsAsFactors = TRUE)
library(dplyr) #filtraggio
target <- c("zione", "ione", "gione")
DatAllomorph <- filter (DatAllomorph, id == 1)
summary(DatAllomorph)
write.table(DatAllomorph, file="DatAllomorph.txt", sep="\t", row.names=FALSE, quote=FALSE)

m_2 <- gam (anticip.BMU ~ dist.to.MB * allomorph + s(repetition.id, bs = "re"), data=DatAllomorph)
summary(m_2)

#
summary (Dat100$frequency)

Dat100$gruppiF <- cut (Dat100$frequency, c( 1, 10, 23, 53, 100, 1000),
labels=c("F. basse", "F. medio-basse", "F. medie", "F. medio-alte", "F. alte"))

ggplot(Dat100, aes(dist.to.MB, anticip.BMU, group=gruppiF))+ geom_smooth(method= "loess", se = FALSE,
aes(linetype=gruppiF, color=gruppiF), size=1.2)+ coord_cartesian(ylim=c(0, 3))+
ylab("Predizione")+ xlab("distanza da confine di morfema")+ theme_bw()+
facet_wrap (~ conj) +
theme(text = element_text (size=20, face="bold"), legend.title = element_blank()) +
ggtitle(" ") + theme(plot.title = element_text (lineheight=0.8, face="bold",
hjust = 0.5, vjust= 0.5), legend.title = element_blank(),
legend.text=element_text(size=16), axis.text=element_text(size=14))

ggplot(datIntermedie, aes(dist.to.MB, anticip.BMU, group=gruppiF))+
geom_smooth(method= "loess", se = FALSE,
aes(linetype=gruppiF, color=gruppiF), size=1.2)+ coord_cartesian(ylim=c(0, 3))+
ylab("Predizione")+ xlab("distanza da confine di morfema")+ theme_bw()+
facet_wrap (~ conj) +
theme(text = element_text (size=20, face="bold"), legend.title = element_blank()) +
ggtitle(" ") + theme(plot.title = element_text (lineheight=0.8, face="bold",
hjust = 0.5, vjust= 0.5), legend.title = element_blank(),
legend.text=element_text(size=16), axis.text=element_text(size=14))

##### ESPERIMENTO 3: famiglia #####

load("DER_FAM_100.txt")
dat1<-read.delim("DER_FAM_100.txt", sep="\t", quote="", dec=".", stringsAsFactors = TRUE)
dat1$lunghezza <-rowSums(dat1[, c("prefix", "stem", "suffix")])

ggplot(datEpoca, aes(dist.to.MB, anticip.BMU, group=conj))+ geom_smooth(method= "loess", se = FALSE,
aes(linetype=conj, color=conj), size=1.2)+ coord_cartesian(ylim=c(0, 3))+
ylab("Predizione")+ xlab("distanza da confine di morfema")+ theme_bw()+
theme(text = element_text (size=20, face="bold"), legend.title = element_blank()) +
ggtitle("Epoca finale")+ theme(plot.title = element_text (lineheight=.8, face="bold",
hjust = 0.5, vjust= 0.5), legend.title = element_blank(),
legend.text=element_text(size=16), axis.text=element_text(size=14))

#modelli non lineare gam
m1 <- gam (anticip.BMU ~ dist.to.MB * conj + s(repetition.id, bs = "re"), data= datFam)
summary(m1)
m2 <- gam (anticip.BMU ~ dist.to.MB*conj + dist.to.MB * lunghezza +
s(repetition.id, bs = "re"), data= datFam)
summary(m2)

```