

UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Psicologia generale  
Corso di laurea in Scienze Psicologiche Cognitive e Psicobiologiche

Elaborato finale

---

Un paradigma EEG di entrainment neurale per lo studio di strutture linguistiche gerarchiche

---

Anno accademico: 2021/2022



RELATORE: PROF. VESPIGNANI FRANCESCO

LAUREANDA: SAFA YASSIN  
MATRICOLA: 1231690

## Indice

Abstract.....	2
Introduzione .....	3
Capitolo 1: La letteratura: .....	4
Il cervello è sensibile agli indizi ritmici del linguaggio?.....	5
Gli aspetti chiave dello studio di Ding et al. (2016) .....	6
Una replica dell'effetto utilizzando l'EEG, Ding et al. (2017).....	8
Alcune repliche e variazioni del paradigma .....	10
La critica di Frank et al. (2018).....	10
Il nostro studio .....	13
Capitolo 2: Una replica in lingua italiana .....	14
Capitolo 3: Il Metodo.....	16
3.1 Modalità di presentazione del materiale.....	16
3.2 La scelta del materiale linguistico .....	18
3.3 La composizione delle frasi .....	21
3.3.1 Le frasi copulari.....	21
3.3.2 Le frasi miste .....	23
3.3.3 La condizione di controllo .....	24
3.4 Struttura dell'esperimento e compito .....	24
3.5 Caratteristiche fisiche dello stimolo .....	25
Conclusione.....	27
Ringraziamenti .....	28
Riferimenti bibliografici .....	28

## Abstract

Anche se la comprensione del linguaggio è apparentemente svolta senza sforzo esplicito, si è mostrato come questa sia il prodotto di un sofisticato sistema di interconnessioni neurali. Si pensa che il nostro cervello sia in grado di costruire meticolosamente strutture gerarchiche sintattiche e semantiche a partire dal semplice flusso sonoro. Per entrainment a strutture linguistiche gerarchiche si intende il fatto che il cervello, partendo dall'analisi e rappresentazioni degli elementi di base del parlato come sillabe e parole, li combina incrementalmente in strutture sintattiche più ampie e complesse quali sintagmi e frasi. In questa tesi riporto un paradigma per una replica concettuale di uno studio di Ding et al. (2017) che ha studiato, utilizzando l'elettroencefalografia (EEG), la sincronizzazione dell'attività corticale in relazione ad una presentazione ritmica di queste unità linguistiche (sillabe, sintagmi e frasi) aggiungendo una condizione sperimentale che mira a chiarire l'interpretazione dei picchi di criticata da Frank et al. (2018). L'adattamento del paradigma alla lingua italiana ha richiesto di affrontare specifiche problematiche, offrendo al contempo l'opportunità di costruire un disegno sperimentale più pulito e complesso.

## Introduzione

Lei è seduta in un bar, e sente due persone accanto che stanno parlando in una lingua straniera, si sforza, disperatamente, per cogliere di cosa stanno parlando, ma il suono le sembra un mero flusso di onde sonore. Magari riesce a capire se fosse arabo o cinese ma non sarà in grado di capire quali sono le parole che compongono tale flusso, né quando una frase finisca è un'altra inizi. Alcune persone troveranno situazioni come questa, che capitano facilmente, interessanti bizzarre, o divertenti, contemplando la bellezza della diversità del linguaggio umano e la sua complessità. Altri, invece, si sforzano a cercare i bordi delle parole, infastidendosi dal fatto che non riescano a farlo come lo fanno con la loro madrelingua, ogni giorno.

Quindi, ci interessa capire come facciamo a rendere tale flusso di fonemi di una lingua che conosciamo, in una costruzione linguistica segmentata e strutturata. Come facciamo a connettere le caratteristiche fonologiche in modo sistematico, estrarre le caratteristiche sintattiche e semantiche in modo talmente automatico, rapido e affascinante?

In questo studio, cerchiamo di approfondire tali domande e capire cosa succede, non solo a livello linguistico ma anche a livello neurale del sistema nervoso centrale. Usiamo l'EEG per poter studiare le risposte neurali a un flusso di frasi continuo. Ci interessa indagare il fenomeno dell'entrainment che corrisponde a un tipo di sincronizzazione dell'EEG spontaneo che produce dei picchi di intensità a frequenze corrispondenti con le frequenze di presentazione delle unità linguistiche basilari, le sillabe, i sintagmi, e le frasi.

Ci chiederemo se abbiamo una elaborazione del materiale linguistico in entrata, a livello centrale, in modo gerarchico o nettamente seriale, e quali sono le limitazioni. Ciò apre una finestrina alla mente umana e potrebbe aiutare a chiarire la natura di alcuni tipi di difficoltà nell'acquisizione o l'utilizzo del linguaggio in particolari popolazioni.

Tale studio richiede di considerare conoscenze provenienti da differenti discipline come la linguistica, la psicologia e le neuroscienze. L'ampia interdisciplinarietà è un aspetto specificamente interessante dello studio del linguaggio rispetto ad altri argomenti in psicologia sperimentale.

## Capitolo 1: La letteratura:

Le rappresentazioni sintattiche gerarchiche sono indispensabili per la comprensione linguistica? sono rintracciabili a livello neurale? L'attività elettrofisiologica del cervello è sincronizzata con le frequenze del parlato?

La comprensione del linguaggio dipende dalla nostra capacità di costruire strutture sintattiche semanticamente interpretabili, da una sequenza di parole in ingresso (Martorell, 2020). Se presentiamo un flusso di parole che non permette la costruzione di una struttura sintattica grammaticalmente corretta come “il mangio lei sera che donna”, non è facile derivare un significato dell'espressione. Infatti, sebbene uno sappia il significato di ogni parola di tale sequenza lessicale, fa fatica ad estrarre un significato (cioè, comprendere) questa frase priva di strutture sintattiche rilevanti.

Dunque, l'importanza delle informazioni sintattiche nella comprensione linguistica è una evidenza condivisa, tuttavia, il modo nel quale la mente costruisce tali rappresentazioni è dibattuta fra gli studiosi; come avviene tale elaborazione sintattica durante la presentazione del materiale linguistico in ingresso, segue un'analisi sequenziale o gerarchica? Quale ruolo ha frequenza con la quale strutture analoghe vengono elaborate? e se si parla di strutture gerarchiche, a quale livello di astrattezza risiedono, sono rappresentabili mentalmente (psicologicamente) e in quale modo? Sono tracciate da specifiche attività a livello neurale?

Nell'articolo di Ding et al. (2016) viene ipotizzata un'analisi gerarchica delle strutture sintattiche delle frasi: *“To understand connected speech, listeners **must** construct a hierarchy of linguistic structures of different sizes, including syllables, words, phrases, and sentences.”* Ovvero, gli autori ritengono che la rappresentazione gerarchica delle strutture sintattiche (parola, sintagma e frase) sia prerequisito indispensabile per la comprensione. Nel loro esperimento, cercano di tracciare la attività corticale sottostante, usando il MEG (Magnetoencefalografia) e il ECoG (elettrocorticogramma), assumendo che la costruzione di strutture gerarchiche debba rendersi evidente in termini di picchi di sincronizzazione ai vari livelli delle strutture sintattiche astratte. L'argomento è che tale sincronizzazione (definita entrainment neurale) possa essere interpretata nei termini di una attività correlata ad una elaborazione cognitiva che computa i differenti livelli della struttura gerarchica.

Descriveremo tale studio e un ulteriore studio dello stesso gruppo di ricerca (Ding et al. 2017) che ha mostrato come gli stessi picchi possano essere rilevati anche con una tecnica più economica (rispetto a MEG) e meno invasiva (rispetto a ECoG), ovvero utilizzando l'EEG.

Nominiamo alcuni degli studi che hanno mostrato la robustezza dell'effetto dell'entrainment osservato e la sua replicabilità in lingue differenti e con differenti metodi di presentazione del materiale linguistico, ci fermiamo sullo studio di (Martorell, 2021), di particolare interesse per noi, che ha replicato l'effetto in modalità visiva.

A seguire introdurremo la critica di Frank et al. (2018) sulla interpretazione del dato sperimentale ricavato dai due studi che pone dei dubbi sull'attribuzione dei picchi ad un'analisi gerarchica, suggerendo che tale correlato neurale potrebbe altresì emergere all'interno di un'analisi sequenziale, sensibile alla ripetizione temporale di elementi di natura differente dal punto di vista lessicale e distribuzionale (probabilità della presenza in differenti sequenze lineari).

### Il cervello è sensibile agli indizi ritmici del linguaggio?

Torniamo un po' indietro, uno dei temi classici di interesse della psicolinguistica è la ritmicità e l'analisi ritmica del parlato, sia nell'acquisizione del linguaggio (Mehler, 1998, 2017, Goswami, 2019) sia nella comprensione del linguaggio in generale (Davis, 2012), o nell'apprendimento di una seconda lingua (Polyanskaya, 2015).

Il parlato stesso è infatti caratterizzato da alternanze più o meno regolare di elementi (p.e. sillabe) forti e deboli dal punto di vista acustico e percettivo (ritmo) ove forte e debole corrisponde a durata e intensità di elemento sonori (principalmente vocali). Anche se abbiamo visto che nelle differenti lingue le persone parlano in maniera molto diversa, con velocità, e accenti diversi, alcuni autori (e.g. Poeppel, 2020) suggeriscono che la frequenza del parlato è abbastanza regolare fra lingue e individui con picchi di frequenze uguali a 4-5 Hz e che il parlato è quasi-ritmico.

Il parlato stesso ha caratteristiche ritmiche a prescindere dalla lingua e la condizione di uso, ci si può quindi chiedere se il nostro cervello sia sensibile a tale ritmicità temporale e la sfrutti per combinare gli elementi linguistici in ingresso in parole, sintagmi e frasi, e se sì, come possiamo scoprirlo?

Le oscillazioni rilevabili con l'EEG, la MEG o EcoG, sono tipicamente distinte in bande (intervalli) di frequenza che corrispondono alla somma di attività di ampie popolazione di neuroni sincronizzate a livello spaziale e temporale di modo da emergere a livello di oscillazioni registrabili tramite elettrodi macroscopici, molte ricerche hanno cercato di associare all'aumento o la diminuzione dell'intensità di segnale nelle varie bande a specifici

processi mentali legati allo svolgimento di differenti funzioni psicologiche come ad esempio attenzione, memoria e consapevolezza (Ward, 2003).

Le tecniche di elettrofisiologia (MEG, EEG e ECoG) mettono in luce variazioni nelle dinamiche di oscillazioni di ampie popolazioni neurali che possono essere empiricamente messe in relazione con specifici processi cognitivi, e in questo caso con i processi di segmentazione e di organizzazione gerarchica del parlato.

In effetti, molti studi hanno usato tali tecniche per suggerire una sincronizzazione corticale alle caratteristiche acustiche del parlato, ma principalmente al livello delle sillabe. (si veda ad esempio Poeppel, 2019, Kraus, 2014).

[Gli aspetti chiave dello studio di Ding et al. \(2016\)](#)

Negli studi di Ding et al. (2016, 2017) hanno affrontato queste due questioni, primo, c'è una sincronizzazione corticale con il parlato udito? Secondo, è possibile tracciare una sincronizzazione non solo al livello delle sillabe ma anche a livelli sintatticamente più elevate e astratte (sintagma, frase) e in conseguenza, supportare un'analisi gerarchica del parlato?

L'ipotesi principale degli autori è stata: in un flusso di sequenze di parole, non ci sono bordi chiari o indizi acustici fissi riguardo l'inizio di un'unità linguistica gerarchica (parola, sintagma e frase), quindi sicuramente tali strutture sintattiche sono elaborate e analizzate internamente e in modo incrementale e online durante la comprensione. È vero che limiti di parole e aspetti strutturali correlano con elementi acustici, come ritmo e prosodia (Mehler, 2002) o con aspetti statistici legati alle frequenze delle probabilità di transizione fra sillabe e fonemi (Saffran, 1996), tuttavia, gli autori ritengono che ci siano altri fattori in gioco, legati a una modulazione temporale e neurale a livelli anche astratti (oltre la sillaba) e che consente l'analisi complessa, combinatoria e grammar-dependent del parlato in ingresso sulla base del riconoscimento di elementi lessicali, anche deprivando lo stimolo degli indizi acustici che correlano con rappresentazioni gerarchiche.

In Ding (2016), sono stati condotti sette esperimenti, dove hanno presentato, acusticamente, delle sequenze di parole monosillabiche in diverse condizioni sperimentali e in differenti lingue (cinese, inglese) dove i vari livelli di strutture gerarchiche sono presenti.

Le parole erano presentate a frequenza fissa di 4 Hz, e nella condizione sperimentale queste componevano delle frasi composte da due sintagmi (sintagmi a frequenza 2Hz e frasi a 1Hz) Come nell'esempio (1).

(1) [[new]<sub>T=250ms</sub> [plans]]<sub>T=500ms</sub> [[gave] [hope]]<sub>T=1sec</sub>

L'ipotesi è che la presentazione sequenziale di frasi come in (1) dovrebbe mettere in luce dei picchi di attività elettrica cerebrale (picchi di entrainment) non solo alla frequenza di stimolazione sillabica (4Hz) ma anche alla frequenza dei sintagmi (2Hz) e delle frasi (1Hz). Al fine di poter mettere questi eventuali picchi in relazione a processi di costruzione di raggruppamenti gerarchicamente organizzati è necessario evitare che nel segnale acustico ci siano delle proprietà che variano alle frequenze di interesse. Sia il ritmo (alternanza di sillabe forti e deboli in termini di durata e intensità) che la prosodia (variazioni della frequenza fondamentale tipicamente calante all'interno di un enunciato) sono aspetti del parlato che correlano fortemente con strutture gerarchiche, al fine di escludere che eventuali picchi di entrainment possano riflettere aspetti percettivi relativi a tali variabili oltre a normalizzare la durata di ogni sillaba (a 250ms) il materiale acustico è stato sintetizzato sillaba per sillaba in modo da non presentare regolarità ritmiche o prosodiche.

Durante l'ascolto è stata misurata l'attività corticale tramite Magnetoencefalografia (MEG), e l'attività elettrica generata dagli stimoli sperimentali (sequenze di frasi quadrisillabiche) è stata confrontata in differenti esperimenti con differenti tipi di condizioni di controllo nelle quali non erano presenti strutture gerarchiche linguistiche (*si veda capitolo 3.2*)

I risultati hanno permesso di mettere in luce, come ipotizzato, la presenza di picchi di potenza (*power spectrum*) nell'analisi spettrale dell'attività evocata media del periodo di stimolazione alle frequenze corrispondenti alla frequenza di stimolazione (sillaba/parola) e ai due livelli delle unità linguistiche gerarchiche (sintagma e frase), mentre nelle condizioni di controllo emergeva unicamente il picco relativo alla frequenza di stimolazione (sillaba).

Una delle condizioni di controllo consisteva in frasi di una lingua non conosciuta dai partecipanti e anche in questo caso i picchi sono presenti solo quando i soggetti sono esposti a una lingua che comprendono, quando i partecipanti che non conoscono il cinese ascoltano frasi in cinese, si è visto solo una sincronizzazione neurale al livello di sillaba.

Sulla base di tali evidenze Ding et al (2016) hanno concluso che tale sincronizzazione bilaterale non è altro che il risultato di un processo di costruzione incrementale interno e gerarchico di strutture sintattiche, coinvolto nella comprensione frasale, avendo escluso in modo chiaro e replicato che tali picchi riflettono aspetti non controllati di periodicità all'interno dello stimolo acustico stesso.



Una replica dell'effetto utilizzando l'EEG, Ding et al. (2017)

Ding et al. (2017) hanno replicato uno degli esperimenti del precedente studio però con la misurazione via l'elettroencefalogramma (EEG), una tecnica più disponibile e accessibile. I risultati confermano il precedente effetto e la presenza di sincronizzazione al livello delle sillabe/parole, sintagmi e frasi.

Il materiale, in lingua inglese, è preso dal sesto esperimento di Ding et al, (2016), e composto come spiegato sopra da quattro parole monosillabiche che formano una frase del tipo: [[XX]NP [XX]VP] e la stimolazione critica consiste nella presentazione sequenziale senza pause di una sequenza di 12 frasi di questo tipo per ogni trial. In questo studio, la durata della sillaba/parola è  $T=320\text{ms}$ .

L'outlier della condizione sperimentale consisteva nel sostituire tre parole dalle frasi normali con tre parole random.

La condizione di controllo consisteva nel presentare uno stimolo acustico sintetizzato ricombinando le sillabe utilizzate nel materiale sperimentale con la tecnica dello slicing. Questa tecnica consiste nel tagliare le frasi acustiche in cinque parti (ognuna  $T=72\text{ms}$ ) e rimescolare una parte (quinto della frase) tra le frasi (conservando la sua posizione), il risultato si consiste in un parlato nel quale sono conservate le sue proprietà acustiche generali ma non è intelligibile. Nella condizione di controllo ci sono anche otto frasi outlier, nelle quali erano inserite quattro parole in inglese.

Al soggetto vengono presentati 30 trial della condizione frasale (22 normali, 8 outlier) e 30 trial della condizione di controllo (22 normali, 8 outlier), in ordine bilanciato tra i soggetti. I partecipanti avevano un semplice compito di dedizione di outlier alla fine del trial, premendo un tasto.

Nell'analisi di dati, sono stati estratti differenti indici: la potenza evocata (*evoked power spectrum*), che riflette la potenza della risposta EEG sincronizzata con il parlato in ingresso, la potenza indotta (*induced power spectrum*) che misura la potenza delle risposte EEG non sincronizzate con il parlato e la coerenza di fase fra i trials (*inter-trial phase coherence*) che misura il grado di phase-locking tra i singoli trial. Ognuna di queste misure è stata estratta per l'intero range di frequenza di interesse (fra 1/1.28 e 1/0.32 Hz). È importante sottolineare la differenza tra la potenza evocata e indotta. Tutti e due sono componenti diverse delle risposte neurali che potrebbero essere il risultato di un cambiamento endogeno (p.e. pensiero) o un cambiamento esogeno (p.e. stimolo), tuttavia la loro differenza resta in quanto siano phase-

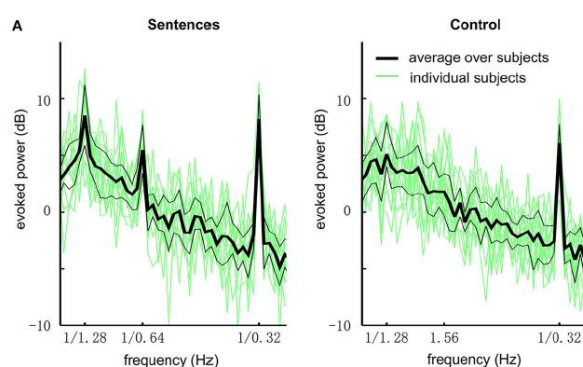
locked (PL) con un event (Chen, 2012). Le risposte evocate sono ad aggancio di fase (PL) con lo stimolo presentato mentre le risposte indotte variano tra i trial ci si aspetta che si azzerino facendo la media lineare delle risposte. la differenza funzionale tra le due risposte è stata attribuita a una modulazione Top-down nel caso delle risposte indotte, e modulazione Bottom-up nelle risposte evocate (Tallon-Baudry and Bertrand, 1999).

A livello statistico l'ipotesi nulla riguardava l'assenza di una differenza significativa della potenza normalizzata del segnale  $En(f)$  tra le frequenze target (di sillaba, sintagma e frase) e le frequenze attorno. Mentre l'ipotesi nulla per i test sulla coerenza di fase inter-trial era l'assenza di una sincronizzazione significativa con lo stimolo in ingresso, ovvero una distribuzione uniforme della alternanza di fase  $\theta_k$ .

Per l'analisi della potenza, è stato effettuato un test di potenza globale, dove la banda di potenza è calcolata per ogni elettrodo per poi calcolare la media di tutti gli elettrodi. facendo la media tra i soggetti si sono rilevati picchi di potenza nelle frequenze target:  $F_{frase}$  (1/1.28Hz)  $F_{sintagma}$  (1/0.64Hz) e  $F_{sillaba}$  (1/0.32Hz), che porta a rifiutare l'ipotesi nulla e concludere che ci sia una differenza effettiva di potenza di segnale in queste annotate frequenze. Nel caso della condizione di controllo, c'era solo una differenza significativa alla frequenza sillabica. (Figura 1)

Analogamente, anche l'analisi della coerenza di fase è stata condotta sulla media tra i canali e si sono trovati tre picchi corrispondenti ad ogni livello di struttura gerarchica per la condizione sperimentale ed il solo picco per la condizione di controllo.

FIGURA 1 (A). La potenza evocata delle risposte EEG. La linea nera grossa segna la media dei soggetti, le linee verdi segnano i dati dei singoli soggetti. Le risposte alle frasi presentano tre picchi ai livelli di frase (1/1.28Hz) di sintagma (1/0.64Hz) e di sillaba (1/0.32Hz), in contrasto nella condizione di controllo le risposte presentano un solo picco al livello della sillaba. Presa da Ding et.al, 2007, *Frontiers in human neuroscience*, 11, p.5.



Il test della potenza indotta (delle risposte non-phase locked) e che si ottiene tramite la sottrazione della potenza evocata dalla media della potenza, non ha evidenziato nessun picco di potenza nelle frequenze di interesse.

I risultati supportano l'ipotesi degli autori, che ci sia una specifica attività neurale nelle frequenze target, e siccome lo stimolo non includeva indizi ritmici o prosodici, tali picchi

sono interpretati in termini di un'analisi incrementale (*phase-locked*) e gerarchica che permette al soggetto di segmentare e aggregare il materiale sonoro in unità linguistiche complesse.

#### Alcune repliche e variazioni del paradigma

Ulteriori studi hanno confermato la presenza dell'effetto di entrainment osservato in diverse lingue. Uno studio cercava di individuare spazialmente le regioni cerebrali responsabili della costruzione di differenti unità gerarchiche (Sheng, 2019), un altro studio ha cercato la relazione tra la prosodia (implicita e esplicita) e l'analisi sintattica gerarchica in tedesco (Glushko et al., 2020), ancora un altro studio ha studiato l'effetto dell'attenzione a tale modulazione temporale con le diverse unità gerarchiche, in inglese (Sokoliuk et al. 2021). Inoltre, uno studio, fatto in spagnolo e in basco, ha replicato l'effetto usando degli stimoli visivi. Introducendo una nuova metodologia dello studio dell'entrainment che verrà adattata anche da noi:

#### Lo studio di Martorell (2021)

lo studio sul “tracciamento neurale della sintassi da una prospettiva cross-linguistica” presentato nella conferenza della società della neurobiologia del linguaggio (SNL 2021) (Martorell, Simona Mancini, Molinaro & Carreiras 2021) replica l'effetto di Ding et al. (2016) con un paradigma sperimentale simile però in modalità visiva.

A soggetti bilingue (spagnolo e basco) viene misurata l'attività MEG mentre guardano un flusso di frasi fatti in tre condizioni sperimentali: solo sintagmi nominali [NP], verbo + sintagmi verbali [verbo] [NP]<sub>spagnolo</sub> [NP][Verbo]<sub>Basco</sub>, e frasi interi [Aux][Verb][NP]<sub>spagnolo</sub> [NP][Verb][Aux]<sub>Basco</sub>.

I loro risultati chiave riguardano la replicazione dell'effetto dell'entrainment, che traccia anche le differenti strutture della direzionalità della testa dello spagnolo e del basco, iniziali di testa (*head-initial*) e finali di testa (*head-final*) rispettivamente; un'evidenza che supporta l'ipotesi dell'entrainment neurale con le strutture sintattiche. Inoltre, la replicabilità dell'effetto usando stimoli visivi.

#### La critica di Frank et al. (2018)

Frank et al. (2018) dedicano uno studio alla critica dell'interpretazione funzionale dei picchi di frequenza rilevati da Ding et al. (2016, 2017), definendo la loro attribuzione a processi astratti di costruzione di strutture gerarchiche come spericolata e non necessariamente vera relativamente alla natura dei processi neurali coinvolti nella comprensione.

Frank et al. (2018) ritengono che l'insieme dei processi mentali che governano e guidano le capacità di comprensione linguistica sia complesso e non facilmente descrivibile in un modello che distingue funzioni puramente sintattico-gerarchiche da altri tipi di processi che tracciano ad aspetti distribuzionali (probabilità e probabilità di transizione) e semantico-lessicali. Questo porta a criticare la visione di Ding et al. (2016, 2017) che semplifica tale complessità in modo dicotomico; l'analisi incrementale che risiede nel livello delle parole e l'analisi incrementale di natura gerarchica che permette la costruzione dei livelli sintatticamente più "elevati". Per Frank et al (2018) la statistica ha un ruolo non solo al livello delle parole ma anche nel "chunking" e/o e nella costruzione di strutture sintattiche. Frank et al. (2018) non criticano il paradigma e i dati sperimentali ottenuti ma la loro interpretazione. Cioè, non pensano che ci sia una relazione nettamente causale tra l'entrainment neurale e l'elaborazione di unità sintattiche gerarchiche come proposto, bensì i picchi di potenza potrebbero essere il risultato della presenza di una mera regolarità di aspetti semantico-lessicali (proprietà specifiche delle singole parole) e che quindi i risultati sperimentali possano essere ricavati da un'elaborazione lessicale che non va oltre il livello di parola.

Frank et al. (2018) creano un modello computazionale di semantica distribuita, simulando le frasi negli esperimenti di Ding et al.(2016) con vettori numerici che segnano la similarità nelle proprietà semantica e sintattico tra le parole (p.e. parole che appaiono nello stesso contesto i loro vettori assumono valori simili). Tale modello va addestrato usando una rete neurale di tipo feed-forward, con delle unità nascoste e con unità di input e output che corrispondono al tipo di parola. Il modello viene esposto a quantità di data enorme in inglese e in cinese. La simulazione del dato di spettro di frequenza EEG/MEG è svolta nell'assunzione che vettori simili diano luogo ad attività neurali simili in un dato intervallo temporale e debbano emergere dei picchi alle frequenze alle quali parole con vettori simili sono elaborate. Simulando sequenza simili a quelle usate da Ding et al (2016, 2017) gli autori sono stati in grado di mettere in luce picchi di frequenza a livello del sintagma e della frase sulla sola base di caratteristiche semantico-lessicali delle parole. (I pesi cambiano quando due parole si presentano nello stesso contesto, p.e. due verbi, due aggettivi sinonimi)

Gli autori trovano una potenza con una variazione assai simile ai risultati dei soggetti umani, ovvero, il modello predice un picco di potenza alla frequenza di parole (4Hz), di sintagmi (2Hz) e di frasi (1Hz).

Secondo gli autori, siccome il modello non includeva strati di rappresentazioni di sintagmi né di frasi (solo proprietà di parole) allora la spiegazione dei picchi del modello sono dovuti alla ricorrenza di proprietà lessicali. Infatti, nel caso nella condizione sperimentale di Ding et al. (2016), la seconda parola nella sequenza è un nome (un'entità) e la quarta è un verbo (un'azione) e siccome tali parole (apparentemente diverse ma sostanzialmente aventi proprietà lessicali simili) hanno vettori di valori simili, ciò spiega la predizione di picchi nella potenza calcolata attorno le loro frequenze.

Inoltre, dal modello si osserva che tali picchi sono conservati anche quando viene presentata “un’insalata di parole” che consiste in una stessa sequenza di parole  $[[XX]_{NP} [XX]_{VP}]$  però dove le posizioni delle parole funzione sono preservate ma le altre parole vengono sostituite con parole random. Il modello computazionale in questo caso preserva i picchi al livello di frase (1Hz) anche se non si presentano delle frasi (sintatticamente valide), un risultato che non bisogna verificarsi secondo l’ipotesi di Ding et al. (2016). Ragion per cui l’ipotesi va ulteriormente verificata con altre condizioni sperimentali. (Figura 2)

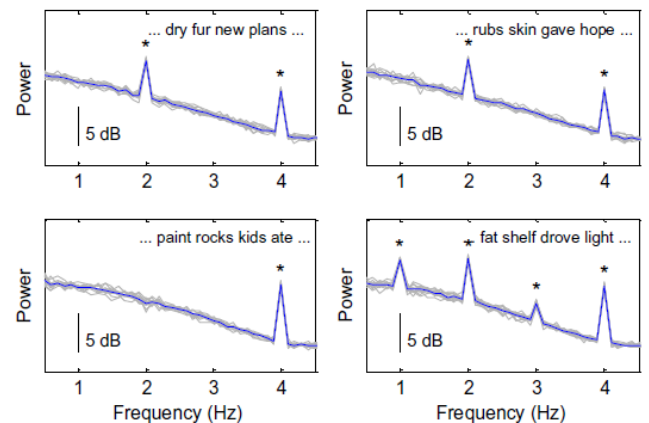


FIGURA 2. Modello di predizione non studiato da Ding et al. (2016). In alto a sinistra, solo sintagmi nominali. In alto a destra, solo sintagmi verbali. In basso a sinistra, un’insalata di parole. In basso a destra, un’insalata di parole con le parole funzione in posizione. Presa da Frank et.al 2018, Language, Cognition and Neuroscience, 33(9), p.1216.

È importante far notare che la critica di Frank et al. (2018) non suppone una assenza di processi gerarchici nella comprensione linguistica, e non esclude la presenza di rappresentazioni mentali di essi, però ci chiede di ripensare il ruolo dell’analisi gerarchica se c’è, dove si verifica e come è rappresentata nel cervello. Soprattutto che si tratta di unità linguistiche di un valore di astrattezza elevato (“sintagma”, “frase”) e che è difficile di pensare che siano favoriti così automaticamente piuttosto di processi più elementari (sequenziali e sensibili alla frequenza di parole).

## Il nostro studio

Il nostro studio sarà una replica concettuale dello studio di Ding et al. (2017), perciò, il primo scopo sarebbe di testare i risultati sperimentali del loro esperimento, usando la stessa tecnica di neuroimaging (EEG) e un paradigma di natura simile (condizioni dove ci sono varie strutture gerarchiche, la parola, il sintagma e la frase) però adattata con stimoli visivi invece di quelli acustici. Ci aspettiamo di avere risultati simili e una differenza significativa tra la condizione di controllo (solo sillabe) e la condizione che contiene strutture gerarchiche (sintagmi, frasi). Il secondo scopo è di testare la critica di Frank et al. (2018), aggiungendo una condizione sperimentale dove le frasi sono di varie strutture grammaticali ma appaiono con una frequenza fissa (e quindi non esisterebbe una ripetizione di un'unica struttura frasale). Tale condizione è cruciale per determinare se effettivamente risultano delle differenze nei picchi di potenza nelle frequenze target legate a un'analisi sintattica gerarchica (la mente conosce i bordi delle frasi anche se sono di strutture diverse), oppure, nel caso che non ci fossero dei picchi ciò potrebbe supportare la critica di Frank et al. (2018) sul fatto che i risultati precedenti fossero legati a una ripetizione di specifiche caratteristiche lessicali.

**Il paradigma:** Nell'esperimento previsto verranno presentati visivamente dei treni di stimoli visivi ad una frequenza di stimolazione fissa (una sillaba ogni 250ms,  $f_s=4\text{Hz}$ ), le condizioni sperimentali consisteranno nella sequenza di 30 trial (22 normali, 8 outlier), ogni trial è composto da 12 frasi.

la frase è composta da 6 sillabe che vanno a comporre una frase composta a sua volta da due sintagmi di 3 sillabe (p.e. [il ca-ne cor-re-va], [il va-so è rot-to], [man-gia-mo sal-sic-cie]) e si dividono in due tipi: frasi copulari, frasi miste (si veda capitolo 3.3). le due condizioni sperimentali saranno confrontate con sequenze casuali di sillabe che non vanno a costituire frasi intellegibili (condizione di controllo).

Il disegno è bloccato, ad ogni soggetto vengono presentati tre blocchi (30 trial) per ogni condizione, le condizioni sono bilanciati tra i partecipanti.

Il compito del/della partecipante sarà di individuare alcune sequenze outlier (catch trials) caratterizzate dalla presenza di frasi semanticamente anomale o con ripetizione di parole che non saranno analizzate. (premendo un tasto alla fine di ogni trial). Verranno valutate le ampiezze di picchi alla frequenza di presentazione di sintagmi ( $f_s/3$ ) e delle frasi ( $f_s/6$ ) delle condizioni sperimentali relativamente alle condizioni di controllo in linea con precedenti studi (Ding et al. 2016, 2017). Durante lo svolgimento del compito verrà misurato in continuo il segnale EEG proveniente da 64 siti encefalici.

## Capitolo 2: Una replica in lingua italiana

Una sfida o un'opportunità?

Nello studio MEG di Ding et al. (2016), gli esperimenti sono stati condotti in cinese e in inglese, l'esperimento EEG di Ding et al. (2017) è stato fatto completamente in inglese. L'adattamento del paradigma ad altre lingue comporta sfide problematiche: inglese e cinese hanno molte parole monosillabiche permettendo di costruire frasi brevi in termini di numero di sillabe. Al contempo le lingue utilizzate da Ding et al. (2016) mostrano anche delle limitazioni legate alla scarsa presenza di morfologia e flessione del fatto che l'ordine delle parole è piuttosto rigido. In questo contesto l'italiano può essere un interessante banco di prova in quanto ha una ricca morfologia flessiva (flessione per genere numero di molti elementi lessicali), ampia variabilità nel numero di sillabe che compongono le parole, relativa libertà nell'ordine delle parole (p.e. soggetto post-verbale, dislocazioni). Queste caratteristiche permettono di guardare al paradigma con nuovi occhi, rispondendo ad alcune domande lasciate aperte e valutando l'estendibilità e la replicabilità dei risultati originali a livello cross-linguistico.

Ci chiediamo anche quali difficoltà emergano da questa scelta, se siano legati a fattori linguistici o applicativi e se quali adattamenti e modifiche al paradigma siano necessari.

Innanzitutto, le frasi che hanno usato Ding et al. (2016, 2017) sono costruiti da parole monosillabiche. Mentre in inglese e cinese ci sono molte parole monosillabiche, in italiano la loro presenza è limitata. Perciò abbiamo scelto le sillabe come i blocchi di base delle frasi. Un problema in quanto il paradigma, i tempi di presentazione e il modo di presentazione vadano ripensati. Tuttavia, in questo caso la sillaba non corrisponde sempre a una parola e ciò potrebbe aprire la porta a creare disegni sperimentali futuri con una metodologia nuova.

Secondo, una caratteristica particolarmente utile che l'italiano ci offre è la flessibilità nell'ordine delle parole in quanto con un numero limitato di sillabe è possibile costruire molte strutture differenti e non solo sequenze di sintagmi nominali e verbali come in Ding (2017) (esempio 2). La possibilità di lavorare con sequenze nelle quali sintagmi e frasi hanno una frequenza fissa mentre elementi lessicali (nomi, verbi) appaiono casualmente in posizioni differenti potrebbe essere alquanto utile per dirimere fra l'interpretazione originale dei picchi e l'interpretazione alternativa suggerita da Frank et al. (2018) in termini di risposta alla presentazione ritmata di specifici tratti semantico-lessicali.

- (2) new plans **gave** hope dry fur **rubs** skin  
 la don.na **can.ta.va** la mo.glie è **ric.ca lo vo.glio** inte.ro  
**man.gia.mo** sal.si.cce ... (si veda capitolo 3)

Una ricchezza morfosintattica offre anche più opzioni e meno vincoli sull'ordine delle parole: se in inglese il soggetto precede necessariamente il verbo, in italiano accanto alla forma non marcata SV è possibile sia omettere il soggetto (soggetto nullo, "arrivo") che avere costruzioni nelle quali il verbo flesso precede il soggetto (p.e. "parte la nave").

- (3) the boy is smart  
 \* is smart the boy  
 il ragazzo è intelligente  
 è intelligente il ragazzo

Terzo, l'italiano è una lingua isosillabica, cioè, una lingua le cui sillabe tendono ad avere una durata costante (Guasti, 2007), questa caratteristica è a nostro favore in quanto lo stimolo è uguale alla sillaba, e dato che la durata della sillaba è tendenzialmente costante, ciò diminuirebbe il bisogno di controllare la differenza nel tempo di lettura di ogni stimolo. (in presentazione acustica di parole monosillabiche si poteva manipolare i tempi di presentazione più facilmente mentre con un materiale visivo la presentazione delle sillabe sarebbe più complicata e meno sistematica se la lingua non fosse isosillabica).

Inoltre, l'italiano è una lingua trasparente, cioè, una lingua in cui ogni grafema corrisponde a un fonema, questo aspetto è rilevante nel nostro paradigma fatto con gli stimoli visivi perché ci dà più controllo sulla grandezza possibile dello stimolo (2-4 lettere). Mentre in inglese una parola monosillabica può avere fino a dieci lettere. (Esempio: *Strengthened*)

In conclusione, l'italiano, per certi aspetti, è molto diverso dall'inglese e il cinese, ciò può effettivamente costituire delle problematiche o aggiungere dei fattori da controllare nel paradigma sperimentale, eppure, non c'è dubbio che offre una gamma di possibili indagini strettamente legate alla questione delle unità linguistiche gerarchiche.



## Capitolo 3: Il Metodo

La costruzione del paradigma:

Come spiegato nel capitolo precedente, la costruzione del paradigma sperimentale in italiano ha implicato di ripensare tanti aspetti dell'esperimento di Ding et al. (2017), in questo capitolo discutiamo alcuni aspetti importanti e quali scelte siano state ritenute più adatte ai nostri obiettivi.

### 3.1 modalità di presentazione del materiale

#### 3.2 La scelta del materiale linguistico

#### 3.3 La composizione delle frasi

##### 3.3.1 frasi copulari

##### 3.3.2 frasi miste

##### 3.3.3 condizione di controllo

#### 3.4 struttura dell'esperimento e compito

#### 3.5 Caratteristiche fisiche dello stimolo

### 3.1 Modalità di presentazione del materiale

Lo studio della comprensione del parlato si concentrava di sulla comprensione uditiva infatti in entrambi gli studi di Ding (2016, 2017) sono stati usati stimoli acustici, ovvero, è stata presentata una sequenza isocrona di sillabe, sintetizzate indipendentemente, sia in inglese che in cinese, senza un gap acustico tra le frasi.

Un altro approccio, implementato da Martorell et al. (2021), consiste nello studiare lo stesso fenomeno di entrainment utilizzando stimoli visivi, cioè una sequenza di parole presentata in modalità visiva al soggetto, ma dove gli stimoli appaiono con una durata costante (200-500ms).

La presentazione di stimoli visivi ad alta velocità nella stessa locazione allo schermo si chiama "Rapid Serial Visual Presentation" (RSVP), un metodo già usato ampiamente per lo studio dell'elaborazione sequenziale del linguaggio. Potter (1984) descrive quanto questo metodo possa essere utile nel testare ipotesi riguardo l'elaborazione delle frasi e dei testi. Potter ritiene che il RSVP sia più simile all'ascolto che non a una vera propria lettura. Ciò è dovuto al fatto che il soggetto non ha nessun controllo sul tempo di presentazione dello stimolo (come nel percepire un suono) e che non possa "tornare indietro" e riguardare gli

stimoli presentati precedentemente (come accade invece nella lettura naturale). Inoltre, siccome gli stimoli vengono presentati in modo seriale e nello stesso posto sullo schermo, non riveste alcuna importanza la dinamica dei movimenti oculari e la visione parafoveale, aspetti fondamentali della percezione del linguaggio in modalità scritta. In confronto con il self-paced reading, al soggetto non è richiesto di interagire con il sistema sperimentale, ciò rende il RSVP più adatto agli esperimenti con misurazioni elettrofisiologiche (p.e. ERPs) dove si cerca di minimizzare gli artefatti oculari e muscolari. Dunque, il RSVP è un metodo dello studio del linguaggio dove invece di misurare il tempo di lettura impiegato dal soggetto, viene misurata l'accuratezza, la latenza del giudizio o la capacità di rievocazione del soggetto in relazione a variabili indipendenti (difficoltà sintattica, tipo di frase), lasciando i tempi di lettura fissati dallo sperimentatore.

Da un punto di vista teorico riteniamo che se i picchi corrispondono a processi di strutturazione gerarchica tali operazioni di alto livello dovrebbero essere indipendenti dalla modalità di presentazione degli stimoli (acustica/parlata vs visiva/ortografica) e infatti Martorell et al. (2021) hanno messo in luce effetti analoghi a quelli di Ding (2017).

In Ding et al (2017) la normalizzazione della durata delle sillabe a 250ms avveniva in modo complesso in quanto sintetizzandole ad una stessa velocità di parlato in alcuni casi la durata era leggermente superiore e in altri casi inferiore, ciò ha portato gli autori a inserire gap di silenzio per le sillabe corte e tagliare parte delle sillabe più lunghe. Il fatto che il RSVP dia la possibilità allo sperimentatore di manipolare più liberamente i tempi di lettura e che in modalità ortografica non è chiaramente necessario privare gli stimoli di aspetti ritmici e prosodici, permettendo di mantenere comunque la natura sequenziale della stimolazione e l'iso-sincronia della presentazione degli stimoli, rende secondo noi la presentazione visiva più semplice e flessibile rispetto alla presentazione acustica che diviene invece fondamentale solo se si attribuisce un ruolo fondamentale nella costruzione di strutture gerarchiche a aspetti prosodici e ritmici. .

È importante far notare che c'erano altri fattori in gioco che ci hanno fatto favorire stimoli visivi piuttosto che stimoli acustici; la facilitazione a livello tecnico legato alla creazione degli stimoli, la possibilità di mettere in confronto lo studio di Martorell et al. (2021) e Ding et al. (2017).

Il fatto che la lingua italiana, a differenza di inglese e cinese, non presenti tante parole monosillabiche permette di pensare a due alternative per implementare il paradigma: mentre

in Ding et al (2017) sillabe e parole coincidono e quindi entrambe sono presentate a ritmo fisso, in italiano si deve scegliere fra presentazione isocrona di parole (formate da una o più sillabe) e presentazione isocrona di sillabe con la conseguenza che i limiti di parole saranno presenti nel flusso a intervalli irregolari. Martorell et al (2021) hanno optato per la prima opzione mentre per noi è più interessante la seconda opzione nell’ottica che se i picchi corrispondono a processi di costruzione di sintagmi e frasi questi dovrebbero presentarsi anche in situazioni nelle quali il ritmo al quale sono presentate parole non sia regolare.

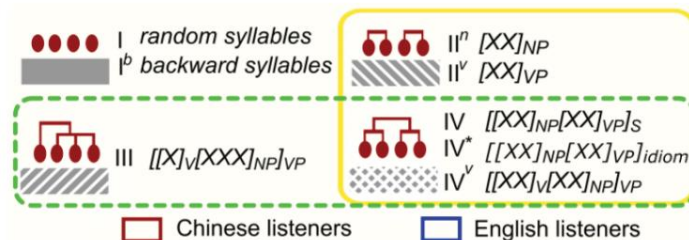
### 3.2 La scelta del materiale linguistico

Le frasi usati nello studio di Ding et al. (2017) sono frasi costruite da quattro parole monosillabiche, sia in inglese che in cinese. Esempio: *fat rat sensed fear*, la struttura variava tra 4 condizioni sperimentali (figura 1).

fig.1:

le quattro condizioni principali usati negli esperimenti di Ding et al. (2016):

1. Sillabe randomizzate [XXXX]
2. Solo sintagmi [XX]<sub>VP/NP</sub> [XX]<sub>VP/NP</sub>
3. Frasi senza sintagmi a frequenza fissa [[X]<sub>V</sub>[XXX]<sub>NP</sub>]
4. Frasi con sintagmi a frequenza fissa (ogni due sillabe) [XX]<sub>NP</sub>[XX]<sub>VP</sub>



Dunque, la variazione tra le quattro condizioni principali riguardava la presenza o l’assenza di ogni livello delle unità linguistiche gerarchiche per poi testare la corrispondente attività corticale. In italiano, come precedentemente accennato, non ci sono tante parole monosillabiche e risulta quindi estremamente complesso creare frasi composte di sole quattro sillabe (p.e. “*so.no con te*”). Usando frasi composte da 6 sillabe, dove ogni 3 sillabe consecutive costituiscono un sintagma (nominale o verbale), è possibile costruire frasi simili a quelle usate da Ding et al (2017), costituite da un sintagma nominale e uno verbale. ad esempio: [la don.na]<sub>NP</sub> [dor.mi.va]<sub>VP</sub>

Inoltre, siccome l’ordine delle parole è maggiormente libera rispetto all’inglese e al cinese, con sei sillabe è possibile costruire differenti strutture con una ampia variabilità della lunghezza e tipo di parole che compongono i due sintagmi trisillabici. (vedi capitolo 3.3.2).

Nel disegnare l’esperimento abbiamo deciso di confrontare implementare una sola condizione di controllo costituita dalle stesse sillabe che verranno usate nelle condizioni sperimentali, randomizzate di modo che non siano presenti parole contenuto di lunghezza bisillabica o

superiore in quanto Ding et al (2016, 2017) hanno ampiamente mostrato che i picchi a frequenza del sintagma e della frase emergono dal confronto di varie condizioni di controllo accomunate dal fatto di non mostrare struttura sintattica.

Tale condizione di controllo verrà contrastata con due condizioni sperimentali, una composta da frasi tutte con la stessa struttura che mostri non solo regolarità di sintagmi e frasi ma anche un'alta regolarità di elementi lessicali e di frequenza di sillabe, per questa condizione si è scelto di utilizzare sequenze di frasi copulari (“Il cane è bello”) per le quali si ha un articolo, un nome, una copula, un aggettivo e ogni 6 sillabe (stessa frequenza della frase) e una parola funzione (articolo o copula) e confine di parole contenuto (nome o aggettivo) ogni 3 sillabe. Una seconda condizione sperimentale presenterà invece sequenze di frasi di differente tipologia che, pur mantenendo la frequenza fissa di frase e sintagma, mostri una ampia variabilità nel tipo e nella regolarità di presentazione di differenti tipi di parole funzione e contenuto. vedi (4)

(4.1) Sillabe randomizzate [lo è ca don il è ne ca bel na la ric]

(4.2) Frasi copulari [[il ca.ne è bel.lo] [la don.na è ric.ca]]

(4.3) Frasi miste [[il ca.ne è bel.lo] [la don.na dor.mi.va]]

Tale disegno sperimentale ci consente per primo, di effettuare separatamente il confronto tra le condizioni (2), (3) rispetto alla condizione di controllo (1). Ci aspettiamo che il confronto tra la condizione (2) e (1) riveli la presenza di picchi alla frequenza di presentazione di sintagmi e frasi, anche se questa manipolazione sperimentale permette di interpretare i picchi anche come risposte a specifici elementi lessicali [Il nome è aggettivo], come suggerito da Frank et al. (2018). Perciò, la condizione (3) è cruciale per chiarire tale ambiguità interpretativa, ovvero, l'introduzione di frasi che hanno diverse strutture grammaticali ci aiuterà a chiarire se la sincronizzazione è effettivamente al livello dei sintagmi e le frasi, evitando il confound dello studio originale (Ding et al. 2017) del fatto che non solo strutture gerarchiche hanno quella frequenza ma anche elementi lessicali (avverbi, nomi, verbi e aggettivi).

Un'ulteriore complicazione nell'adattare il paradigma all'italiano sempre legata alla relativa libertà d'ordine delle parole e dei sintagmi nella frase, deriva dalla possibilità di costruire rappresentazioni frasali con la il secondo sintagma di una frase e il primo sintagma della frase successiva, situazione che denomineremo parsing alternato. Questa situazione è evidente nelle frasi copulari in quanto in italiano è grammaticale sia “*il cane è bello*” che “*è bello il*

*cane*". È vero che nella seconda frase richiede una particolare accentuazione prosodica, legata ad un focus sul predicato ma tale struttura è valida e corretta quanto lo sia la prima. Quindi, all'interno di un flusso di frasi copulari presentato senza pause o indizi prosodici come previsto dal paradigma, si potrebbe avere più di una lettura. Esempio:

...Il cane è bello il tizio è buffo il cuoco è bravo il bimbo...

*Lettura 1:* [il ca.ne è bel.lo] [il tizio è buf.fo] [il cuo.co è bra.vo][Il bim.bo è X.X]

*Lettura 2:* [è XX il ca.ne] [è bel.lo il tizio] [è buf.fo il cuo.co] [è bra.vo Il bim.bo]

Tale alternanza di parsing non è possibile se il soggetto considera correttamente l'inizio della sequenza ma nel caso ci sia una perdita di concentrazione durante il compito è possibile che il parser tenti di volta in volta di costruire strutture diverse. Chiaramente l'eventualità che i partecipanti all'esperimento mostrino alternanza di parsing può dipendere dal livello di concentrazione e dalla congruenza semantica e sintattica delle letture alternative.

Infatti, con la perdita di concentrazione si può pensare ad alternanza di lettura in cui il soggetto sia più sottoposto ad iniziare una lettura dalla metà frase, perché quando egli è meno concentrato, magari si perde qualche sillaba, riprendendo la lettura in modo piuttosto casuale.

Dal punto di vista elettrofisiologico, nell'ipotesi di Ding et al. (2017) che il picco f/3 corrisponda alla costruzione del sintagma e f/6 a quello della frase, questa incertezza e variabilità dovrebbe portare a una diminuzione del picco di frase (f/6) e a un eventuale aumento del picco f/3 in quanto alternatamente le frasi possono finire con frequenze differenti ma sempre alla fine di un sintagma. Chiaramente se i picchi dipendono solo da fattori lessicali (Frank et al. 2018) l'alternanza di parsing non dovrebbe comportare alcun effetto.

Due fattori possono facilitare o meno il parsing alternato: la congruenza semantica e di genere o numero. Variando infatti tali fattori fra due frasi consecutive è possibile manipolare la grammaticalità e sensatezza delle frasi. Prendendo gli esempi (5) il parsing alternato produrrebbe le frasi: "è bello il vaso" in (5.1) grammaticale e sensata, "è rotto il cane" in (5.2) grammaticale ma poco sensata e "è bella il vaso" in (5.3), non grammaticale.

(6.1) ... il cane è **bello il vaso** è rotto ...

(6.2) ... il vaso è **rotto il cane** è bello ...

(6.3) ... la cagna è **bella il vaso** è rotto ...

Quindi la congruenza semantica e sintattica intra-frasale e l'incongruenza semantica interfrasale impedirebbe, oppure renderebbe improbabili alcune letture alternative.

La regolarità o meno di eventuali letture alternate, dovute alla natura della lingua italiana, è cruciale nella fase di costruzione di frasi e sequenze. Sembra che sia un punto di debolezza che va controllato, ma in realtà, è una caratteristica che rende il quadro sperimentale in prospettiva più ricco e che può altresì essere sfruttata per chiarire la natura dei picchi di entrainment all'interno del paradigma. In questo primo esperimento, tuttavia, al fine di limitare il numero di condizioni sperimentali si è cercato di evitare di incorrere nel rischio di parsing alternato, utilizzando sequenze che alternano il genere grammaticale. Chiaramente, manipolazioni dell'accessibilità al parsing alternato potranno essere studiate in successivi esperimenti.

### 3.3 La composizione delle frasi

#### 3.3.1 Le frasi copulari

Per comporre le frasi copulari, è stato chiesto a tre tirocinanti del nostro gruppo di ricerca di scrivere 60 frasi copulari ciascuno che hanno la struttura [[il/la XX] [è YY]] dove il XX è un nome bisillabico e il YY è un aggettivo bisillabico, evitando di ripetere nomi e aggettivi. Nonostante l'ampia disponibilità di nomi e aggettivi bisillabici in italiano comporre un ampio numero di frasi con elementi lessicali diversi ha portato a iniziare a utilizzare parole poco frequenti (vedi [materiali supplementari](#)). Per questo motivo si è deciso di costruire il materiale ripetendo due volte lo stesso nome e lo stesso aggettivo in frasi differenti. Negli esperimenti di Ding et al. (2016, 2017) non è dato sapere quante ripetizioni di una stessa parola fossero presenti nel materiale, Al fine di avere 360 frasi, ho analizzato le 180 frasi composte, controllando e rimuovendo eventuali ripetizioni con parole nuove, e comporre ulteriori 180 frasi rimescolando aggettivi e nomi.

Per avere un buon controllo del materiale e minimizzare il parsing alternato utilizzando il genere grammaticale metà delle frasi erano maschili e metà femminili, evitando l'utilizzo di aggettivi bisgeneri (p.e. "grande") e segnalando prima e seconda ripetizione di nomi e aggettivi per evitare che la stessa parola possa comparire due volte in una stessa sequenza

*Aspetti presi in considerazione durante la fase di composizione:*

1. **La grammaticalità:** evidentemente, le frasi devono essere grammaticalmente corrette, seguendo le regole di agreement di genere, numero, tempo e modo verbale dell'italiano.

**2. Il valore semantico:** oltre alla grammaticalità, nella composizione delle frasi si è prestato attenzione al valore semantico di esse, ovvero quanto una frase suoni bene o male ad un orecchio di una madrelingua italiana. Le frasi sono state valutate da tre parlanti madrelingua italiano portando ad una revisione di alcune frasi giudicate strane (vedi esempio):

[la fra.se è lun.ga] (valutata: buona)

[la fra.se è so.la] (valutata: strana)

**3. l'ambiguità semantica:** una frase potrebbe essere grammaticalmente corretta e potrebbe avere un valore semantico valido però va intesa in due modi diversi da due soggetti diversi. Ciò è dovuto a un'ambiguità lessicale/semantica legata magari ad uso di omonimi, che sono parole che hanno una identica forma grafica/fonetica però hanno due diversi lessemi (sono semanticamente distinte). Esempio:

[la fon.te è tur.ca]

**la fonte:** la vena d'acqua *oppure* le origine/la provenienza di un testo (senso figurativo).

Lo scopo era di ridurre tale ambiguità in modo da avere maggior controllo della ricevibilità individuale delle frasi e diminuire le differenze inter-soggettive nell'interpretazione semantica di esse. Soprattutto che la presentazione lineare non ripetibile, discontinua e veloce degli stimoli non consente al soggetto di avere un modo per disambiguare tali frasi.

Ciò occorre perché la struttura del paradigma stesso inibisce al soggetto di sfruttare alcuni indizi di disambiguazione e in conseguenza ri-leggere la frase e ricostruire il suo senso inteso come succede in una lettura normale.

**4. l'idiomaticità:** durante la composizione si è cercato di evitare di includere idiomi o altre espressioni fisse molto frequenti (p.e. “*il dado è tratto*”), usate altresì in alcuni degli esperimenti di Ding et al (2016) in cinese. Questo constrain è dovuto al fatto che sequenze idiomatiche sono frequenti e tipicamente elaborate in modo differente, accedendo al significato in modo non compositazionale e questo potrebbe rendere questi elementi particolarmente salienti durante la sequenza catturando l'attenzione rispetto ad altri e, con ogni probabilità, diminuire i processi di costruzione sintattica incrementale che sottostanno ai picchi di frequenza secondo l'interpretazione di Ding et al. (2016, 2017)

### 3.3.2 Le frasi miste

Come è stato esplicitato prima, la terza condizione delle frasi miste costituisce una condizione cruciale per controllare l'ipotesi di Ding et al (2016, 2017) o quanto sia la critica di Frank (2018) rilevante. Tuttavia, la scelta di tali frasi non dovrebbe essere affatto casuale, infatti, per poter fare tale confronto dobbiamo porre dei controlli sulle categorie grammaticali scelte.

Similmente alle frasi copulari, per avere una prova preliminare, è stato chiesto a due tirocinanti (studentesse magistrali di linguistica) di comporre e categorizzare *cento* frasi utili al nostro esperimento, ovvero con una struttura grammaticale ma che siano composte di 6 sillabe e due sintagmi ciascuno, composto da 3 sillabe.

Le principali strutture emerse sono:

[la ca.sa crol.la.va]	soggetto-verbo
[sa.li.va la lu.na]	verbo-soggetto
[lo vo.glio in.te.ro]	CLITICO PREVERBALE
[vole.va un do.no]	(S)Verbo-Oggetto
[ve.ni.vo in tre.no]	Verbo-Avverbio
[do.ve.vo pen.sar.ci]	CLITICO POSTVERBALE
[gio.ca.re fa be.ne]	Soggetto-Verbo-avverbio
[la gom.ma è spor.ca]	COPULARE

All'interno di queste categorie si è notato come alcune strutture marcate necessitino di una particolare prosodia per essere legali in italiano, come ad esempio “*a giugno si parte*” è una struttura che è usata tipicamente in modo contrastivo (“*a giugno si parte, non a maggio*”) e tipicamente è accompagnata da un accento prosodico di enfasi sul primo sintagma (“*a giugno*”). Si è ritenuto che sia possibile ottenere una ampia variabilità di struttura anche evitando queste strutture che probabilmente sono più complesse da interpretare all'interno di una presentazione sillaba per sillaba priva di indizi prosodici. Tali frasi caratterizzate inoltre da movimento sintattico verso la periferia sinistra potrebbero suggerire e rendere altamente disponibili soluzioni di parsing alternato (aggiunti come “*a giugno*” sono facilmente attaccabili ad elementi precedenti e successivi). *La composizione delle frasi in questa condizione è ancora in corso.*



### 3.3.3 La condizione di controllo

le sillabe randomizzate erano prese dalle frasi composte per le condizioni (2) e (3) però la presentazione casuale di esse, senza avere nessun'unità linguistica di gerarchia maggiore (parola/ sintagma/ frase) implicava pensare ad alcuni constraints.

1. controllo della presenza di parole prodotte casualmente dall'algoritmo di randomizzazione.

2. non avere due sillabe uguali in sequenza.

### 3.4 Struttura dell'esperimento e compito

In questa tabella, riportiamo il confronto delle scelte relative alla lunghezza del trial, composto da una sequenza ininterrotta di farsi, al numero di trials e se questi siano da presentare con un disegno bloccato (tre condizioni in tre blocchi distinti) o misto (blocchi composti da sequenze di tipo differente).

	<b>Ding 2016</b>	<b>Ding 2017</b>	<b>Martorell 2021</b>	<b>OUR EXPERIMENT</b>
<b>Trial</b>	10 frasi- cinese 12 frasi- inglese	12 frasi	24 chunks*	<b>12 frasi</b>
<b>blocco</b>	Exp1: 30 trials  20 normali 10 outlier	30 trials  22 normali 8 outlier	72 trials  12 trial per condizione (12 x 6)	<b>30 trials</b>  <b>22 normali</b> <b>8 outlier</b>
<b>Numero di frasi per blocco</b>	300 frasi – cinese 360 frasi – inglese	360 frasi	864 chunk-spagnolo 864 chunk-Basco	<b>360 frasi</b>

\*chunk = [NP]/[VP]/[Aux.]

Dunque, manteniamo il disegno bloccato come in Ding et al. (2017), nel quale ogni blocco è composto da 30 trial, ogni trial è composta da 22 frasi normali e 8 frasi outlier e in cui viene presentato ai soggetti i blocchi di ogni condizione sperimentale in modo bilanciato. Il compito adottato da Ding (2017) consisteva nel chiedere al soggetto di individuare i *outlier trials* da quelli *normali* premendo un tasto alla fine del trial. Nel nostro paradigma adottiamo anche tale procedura. Anche se teoricamente, il paradigma può essere virtualmente somministrato senza chiedere al soggetto di fare alcun compito (Ding et al.

2017) gli studi precedenti hanno sempre utilizzato un compito per imporre ai partecipanti un simile livello di attenzione alla stimolazione.

### 3.5 Caratteristiche fisiche dello stimolo

#### 1) Il tempo della presentazione dello stimolo (frequenza)

	Ding 2016	Ding 2017	Martorell 2021	OUR EXPERIMENT
<b>Ts (tempo di sillaba)</b>	250 ms (cinese) 320 ms (inglese)	320 ms	350 ms	<b>250ms</b>
<b>Fs (frequenza di sillaba)</b>	4 Hz 1/0.32 Hz	1/0.32 Hz	1/0.35 Hz	<b>4Hz</b>

Scegliere la frequenza degli stimoli (le sillabe in questo caso) non è stato affatto facile, è vero che volevamo una presentazione non troppo veloce che consente al lettore di costruire le unità linguistiche più complesse, ovvero connettere le unità sintatticamente e rappresentarli semanticamente, cioè comprenderli. Però, la presentazione non deve essere troppo lenta, perché una presentazione lenta risulterebbe in una perdita delle informazioni nella memoria di lavoro, e impedisce di avere il flusso di unità linguistiche che assomiglia al flusso sonoro che vogliamo. Inoltre è noto che il rumore elettrico così come l'attività EEG spontanea aumenta di ampiezza al diminuire della frequenza (teorema di equipartizione dell'energia), inoltre a frequenze molto basse è possibile che comuni artefatti del segnale EEG quali variazioni della conduttanza cutanea o lenti movimenti del capo possano interferire in modo specifico con componenti EEG relativamente lente. Passando da 4 a 6 sillabe, anche mantenendo la durata di presentazione di Ding et al (2017) di 250ms l'atteso picco a livello di frase si sposta verso frequenze più basse ( $f/6$ , pari a  $2/3\text{Hz}$  rispetto a  $f/4$ , pari a 1 Hz).

In Potter (1984), si fa il confronto tra le esperienze soggettive di uno studente universitario medio, alla presentazione di parole in varie frequenze. 6 parole al secondo ( $T_s = 160$  ms) è considerata la frequenza moderata durante la lettura, a 4 parole al secondo ( $T_s = 250$  ms) la lettura va valutata come lenta e noiosa, partendo da 12 parole al secondo in su (35-83 ms), la lettura viene ritenuta sempre più difficile, si riconoscono le parole con  $T_s = 83$  ms ma le informazioni non vengono adeguatamente ritenute, alzando la frequenza, si diminuisce l'impressione soggettiva di poter leggere delle parole, però alcune misure oggettive fanno vedere che anche a frequenze alte (16-28 parole al secondo) è comunque presente qualche

tipo di elaborazione linguistica. Queste osservazioni lasciano spazio per una riduzione del tempo di presentazione, al fine di non spingere il picco di frasi verso frequenze troppo basse dell'EEG dove ci si attende un rapporto segnale-rumore più sfavorevole.

Oltre a questi fattori, nel nostro esperimento in specifico, le sillabe sono gli stimoli presentati nel frame invece delle parole monosillabiche, ciò vuole dire che una lettura lenta non va favorita.

Dopo aver costruito il programma completo su Matlab, sarà utile provare a capire quale potrebbe essere la frequenza idonea, prove preliminari suggeriscono che il tempo di presentazione delle sillabe può variare fra i 100ms e i 250 ms che consente di avere la frequenza della frase fra 0.66 e 1.66 Hz.

### *2) le pause tra parole e frasi:*

In Ding et al. (2017), il trial (sequenza) era composto di un flusso di 12 frasi presentate in modo continuo e isocrono senza una pausa tra esse. Solo alla fine del trial il soggetto deve premere un tasto distinguendo i trial normali dagli outlier. Anche noi adattiamo tali variabili e procedura.

### *3) il numero di parole/lettere per frame:*

Nella rappresentazione di stimoli che cambiano velocemente, è importante capire se ci sia un effetto di mascheramento visivo, sia tra le lettere in preferia e in centro sia tra gli stimoli in sequenza. Ciò anche influisce alla scelta del window frame ottimale e il tipo di lettere preferito (maiuscole o minuscole).

In Cocklin et al. (1984), hanno studiato il window size migliore nel caso di uso del RSVP, e hanno concluso che la media di 12 lettere (con lo spazio tra le lettere incluso) per window è ottimale nell'uso di stimoli rapidi.

Siccome i nostri stimoli sono delle sillabe, le lettere di ciascun stimolo saranno limitate tra le 2-4 lettere (dovuto alla natura trasparente della lingua italiana). Ciò rende il frame più sistematico e non molto variabile molto riducendo preoccupazioni relative alla scelta della dimensione del font e di gradienti percettivi fra visione foveale e periferica ma al contempo aumentando il rischio di mascheramento fra sillabe successive di lunghezza analoga.

Ulteriori scelte che dovranno essere prese sono relative al font e lo script (maiuscole, minuscole) e alla luminosità dello stimolo in relazione allo sfondo. Sulla base di alcune prove

preliminari si è visto che è possibile avere parametri di visualizzazione che permettono una percezione piuttosto buona delle sillabe e delle frasi per presentazioni della durata di circa 200ms, la messa a punto dettagliata di tali parametri sarà svolta a seguito della composizione finale degli stimoli nelle varie condizioni, anche in funzione dell'accuratezza nel compito che sarà svolta sulla scorta di alcuni test comportamentali che verranno svolti prima di procedere alle misurazioni EEG.

## Conclusione

Questo elaborato descrive in maniera dettagliata le procedure della costruzione di un paradigma sperimentale e rende esplicito l'importanza della ricerca teorica, concettuale e comparativa e della costruzione scrupolosa del materiale al fine di avere un paradigma sperimentale ben-definito e idoneo ai nostri scopi teorici.

Seguendo la cronologia degli studi fatti sull'entrainment neurale delle strutture sintattiche nella comprensione del linguaggio è cresciuto l'interesse per il paradigma e l'interpretazione del dato e di conseguenza il bisogno di espandere il tema, di indagare meglio sia i punti di d'accordo che di disaccordo tra gli autori. Per di più, la replica dello studio in italiano, benché implichi di fare tanti aggiustamenti al paradigma, offre un'opportunità di arricchimento sperimentale, rafforza le conclusioni stesse riguardo i processi cognitivi coinvolti nell'analisi sintattica del linguaggio, e verifica la replicabilità e la validità cross-linguistica dell'effetto studiato da Ding et al. (2016,2017).

Il prossimo obiettivo prevede la somministrazione dell'esperimento a soggetti di madrelingua italiana (o che l'hanno appresa prima dei tre anni di età). Ci aspettiamo di tracciare differenze dei picchi di potenza tra le nostre condizioni sperimentali, soprattutto tra le condizioni sperimentali (contenenti frasi) e la condizione di controllo (solo sillabe) e siamo entusiasti a scoprire cosa succede nella terza condizione sperimentale con le frasi che appaiono a frequenza fissa ma che hanno strutture grammaticali differenti; una sincronizzazione dei picchi di potenza con la frequenza di frase (fs/6) o di sintagma (fs/3) consoliderà le ipotesi di Ding et al. (2016, 2017) mentre un'assenza di picchi di potenza significanti attorno tali frequenze irrobustisce la critica di Frank et al. (2018), mette in dubbio alcuni aspetti dell'analisi gerarchica, e ci fa postulare ulteriori domande sulla natura dell'elaborazione sintattica del linguaggio durante la comprensione.

## Ringraziamenti

Ringrazio il professor Vespignani per il supporto immenso, per i commenti preziosi e per le interessanti chiacchiere sull'accademia e sul mondo del linguaggio.

Sono anche grata per aver fatto parte del gruppo "Colab" seguito dai prof Vespignani, Peressotti e Navarette, ho imparato tante cose nuove sulla psicolinguistica.

Ringrazio mia mamma, mie sorelle per le chiamate di sostegno emozionale, e tutta la mia famiglia e i miei amici di cuore per essere semplicemente lì.

E grazie a bambù, il gatto più razzista del mondo, che mi faceva svegliare la mattina per andare in aula studio.

## Riferimenti bibliografici

Arditi, A., & Cho, J. (2007). Letter case and text legibility in normal and low vision. *Vision research*, 47(19), 2499-2505.

Assaneo, M. F., Ripollés, P., Orpella, J., Lin, W. M., de Diego-Balaguer, R., & Poeppel, D. (2019). Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. *Nature neuroscience*, 22(4), 627-632.1

Chen, C. C., Kiebel, S. J., Kilner, J. M., Ward, N. S., Stephan, K. E., Wang, W. J., & Friston, K. J. (2012). A dynamic causal model for evoked and induced responses. *Neuroimage*, 59(1), 340-348.

Cocklin, T. G., Ward, N. J., Chen, H. C., & Juola, J. F. (1984). Factors influencing readability of rapidly presented text segments. *Memory & Cognition*, 12(5), 431-442.

Corazza, A., Lavelli, A., Satta, G., & Zanolì, R. (2004). Analyzing an Italian treebank with state-of-the-art statistical parsers. In *Proceedings of the Third Third Workshop on Treebanks and Linguistic Theories (TLT 2004)* (Vol. 1, p. 155).

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature neuroscience*, 19(1), 158-164.

Ding, N., Melloni, L., Yang, A., Wang, Y., Zhang, W., & Poeppel, D. (2017). Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG). *Frontiers in human neuroscience*, 11, 481.

Dobres, J., Wolfe, B., Chahine, N., & Reimer, B. (2018). The effects of visual crowding, text size, and positional uncertainty on text legibility at a glance. *Applied ergonomics*, 70, 240-246.

Frank, S. L., & Christiansen, M. H. (2018). Hierarchical and sequential processing of language: A response to: Ding, Melloni, Tian, and Poeppel (2017). Rule-based and word-level statistics-based

processing of language: insights from neuroscience. *Language, Cognition and Neuroscience*, 33(9), 1213-1218.

Ghitza, O. (2012). On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. *Frontiers in psychology*, 3, 238.

Goswami, U. (2019). Speech rhythm and language acquisition: an amplitude modulation phase hierarchy perspective. *Annals of the New York Academy of Sciences*, 1453(1), 67-78.

Guasti, M. T. (2007). *L'acquisizione del linguaggio. Un'introduzione*.

Keyser, C., & Perrett, D. I. (2002). Visual masking and RSVP reveal neural competition. *Trends in cognitive sciences*, 6(3), 120-125.

Liang, Franklin Mark. "Parola Hy-phen-a -tion di Com-pu-ter ". Tesi di dottorato, Dipartimento di informatica della Stanford University (en) . Rapporto numero STAN-CS-83-977, agosto 1983. Disponibile su: <https://www.tug.org/docs/liang/liang-thesis-hires.pdf>

Martorell, J., Morucci, P., Mancini, S., & Molinaro, N. (2020). Sentence processing: How words generate syntactic structures in the brain.

Martorell, J., Mancini, S., Molinaro, N., & Carreiras, M. (2021). Neural tracking of syntax from a cross-linguistic perspective Poster Society for the Neurobiology of Language Annual 13th Meeting (SNL 2021) Society for the Neurobiology of Language

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human perception and performance*, 24(3), 756.

Ordin, M., & Polyanskaya, L. (2015). Acquisition of speech rhythm in a second language by learners with rhythmically different native languages. *The Journal of the Acoustical Society of America*, 138(2), 533-544.

Patel, A. D., Iversen, J. R., Bregman, M. R., & Schulz, I. (2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current biology*, 19(10), 827-830.

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in psychology*, 3, 320.

Peña, M., Bonatti, L. L., Nespore, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, 298(5593), 604-607.

Potter, M.C. (1984). Rapid serial visual presentation (RSVP): A method for studying language processing. In D. Kieras & M. Just (Eds.), *New methods in reading comprehension research* (pp. 91-118). Hillsdale, NJ: Erlbaum.

Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature reviews neuroscience*, 21(6), 322-334.

Rubin, G. S., & Turano, K. (1992). Reading without saccadic eye movements. *Vision research*, 32(5), 895-902.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926-1928.

Sheng, J., Zheng, L., Lyu, B., Cen, Z., Qin, L., Tan, L. H., ... & Gao, J. H. (2019). The cortical maps of hierarchical linguistic structures during speech perception. *Cerebral cortex*, 29(8), 3232-3240

Sokoliuk, R., Degano, G., Melloni, L., Noppeney, U., & Cruse, D. (2021). The Influence of Auditory Attention on Rhythmic Speech Tracking: Implications for Studies of Unresponsive Patients. *Frontiers in human neuroscience*, 464.

Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in cognitive sciences*, 3(4), 151-162.

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends in cognitive sciences*, 7(12), 553-559.

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in cognitive sciences*, 15(4), 160-168

Woodruff Carr, K., White-Schwoch, T., Tierney, A. T., Strait, D. L., & Kraus, N. (2014). Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. *Proceedings of the National Academy of Sciences*, 111(40), 14559-14564.

Materiali supplementari:

<https://github.com/safayassino/syllibus>