



UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI STUDI LINGUISTICI E LETTERARI
CORSO DI LAUREA TRIENNALE INTERCLASSE IN
LINGUE, LETTERATURE E MEDIAZIONE CULTURALE (LTLLM)
CLASSE LT-12

IL PROCESSO DI ACQUISIZIONE DEL LINGUAGGIO
Analisi dell'*ellissi* nel dialogo caregiver-bambino

Relatore

PROF.SSA EMANUELA SANFELICI

Laureando

VERONICA MAGAGNA
MATR. 1222943 / LTLLM

13 OTTOBRE 2022
ANNO ACCADEMICO 2021/2022

Vorrei dedicare questa tesi....

Ai miei genitori Stefano e Graziella, esempi di vita che mi hanno sempre incoraggiata e sostenuta.

A mio fratello Nicola e al mio fidanzato Eugenio per tutti i momenti di sostegno, divertimento e spensieratezza che mi hanno regalato.

Alle mie nonne, ai miei parenti vicini e lontani che mi hanno dimostrato sempre affetto e stima.

Alle amiche e agli amici partecipi delle mie avventure e disavventure, che hanno percorso con me questa strada fatta di molte sorprese, qualche difficoltà e incredibili soddisfazioni.

Alla mia relatrice e ai professori incontrati nella mia carriera di studentessa, che con la loro professionalità, esperienza e originalità hanno lasciato in me un segno indelebile.

A me stessa, ai miei sogni e al mio futuro.

INDICE

1	Introduction	1
2	Cos'è il linguaggio	3
2.1	Introduzione	3
2.2	Definizione di linguaggio	4
2.3	Differenza tra lingua, linguaggio e comunicazione	5
2.4	Tratti distintivi del linguaggio umano	6
3	L'acquisizione del linguaggio	9
3.1	Come avviene il processo di acquisizione del linguaggio	9
3.2	Principali teorie dell'acquisizione del linguaggio	10
3.2.1	La proposta comportamentista	10
3.2.2	La teoria cognitivo-funzionalista	11
3.2.3	La teoria emergentista	12
3.2.4	La teoria innatista	12
4	Fenomeno linguistico dell'ellissi	15
4.1	Cos'è l'ellissi	15
4.2	Rassegna dei risultati di studi precedenti	16
4.2.1	<i>The acquisition of nominal ellipsis in Greek</i>	16
4.2.2	<i>Acquisition of noun ellipsis in French and in Dutch - consequences for linguistic theory</i>	17
5	Analisi dei dati	19
5.1	La banca dati CHILDES	19
5.2	Le fasi dell'analisi dei dati sulla base di CHILDES Italian Tonelli Corpus	20
5.3	Descrizione per qualità e quantità delle produzioni input e output	20
5.4	Confronto dei dati input e output	25
6	Conclusion	29
7	Summary	31
	Bibliografia	33

1

INTRODUCTION

This dissertation rises from a personal interest and curiosity towards the language acquisition process. It gives general notions on this complex process, on the production of ellipsis in caregiver-child dialogues. How do children learn to speak? This question brings with it countless questions such as: why is it so easy for children to learn to speak? Do children from an early age begin to use certain syntactic structures due to imitation of the caregiver? There have been many attempts to answer all these questions, but at the moment the issue is unresolved and there are numerous theories about it. This work is divided into two parts: the first, consisting of three chapters, is dedicated to theoretical notions on the acquisition of language and ellipsis and the second one, composed by one chapter, is entirely reserved to the analysis of the data gathered from CHILDES Italian Tonelli Corpus.

The aim of this dissertation is first to examine what language is and analyze its distinctive features. Following this first introductory chapter, the second one will focus on the process of language acquisition, that is the process by which humans acquire the capacity to perceive and comprehend language, as well as to produce and use words and sentences to communicate. In this second chapter the main theories of language acquisition will be explained, such as the behavioural and cognitive approaches and the emergent and innate theories. The third chapter will be dedicated to a review of the results of previous studies concerning the spontaneous ellipsis of children. Finally, the different types of *remnant* of ellipsis used by caregivers and children in their conversations during game sessions will be analyzed and discussed in the last experimental chapter. The collected data derive from the transcripts contained within the CHILDES Italian Tonelli Corpus, structured in three sub-corpora corresponding to three children Elisa, Gregorio and Marco.

The motivations that have driven me to deepen this topic are various. First of all, I would like to thank Emanuela Sanfelici, my professor and rapporteur, whose teaching immediately aroused my curiosity and stimulated my desire to explore the theme in question. Carrying out a study on the acquisition of language and in particular on the use of the ellipsis in caregiver-child dialogues is not simple: its complexity implies the existence of several interpretative models, and this is why it is still very difficult to get a conclusion from the current state of knowledge. But it is precisely the vastness and intricacy of the topic with a focus on language with

children that has fascinated me the most. Important in choosing the theme was also the university course Developmental Psychology during which the growth of the child was deepened, considering the different components of social, emotional, cognitive and even linguistic development.

This elaborate was written to offer a theoretical review on language and its acquisition process and a practical one on the use of ellipsis by parents and children during their dialogues.

2

COS'È IL LINGUAGGIO

«Il linguaggio è una forma della ragione umana, con una sua logica interna della quale gli uomini non conoscono nulla.»

Claude Lévi Strauss

2.1 Introduzione

Il presente lavoro intende analizzare il processo di acquisizione del linguaggio, testando il ruolo degli input a cui i bambini sono esposti nell'acquisizione della sintassi. Si tratta di un processo molto complesso, in quanto è caratterizzato da grandissime variabili individuali che riguardano non solo i tempi, ma anche i modi e le strategie di apprendimento.

In passato le ricerche sul linguaggio in età evolutiva cercavano di mettere in evidenza le tappe universali dello sviluppo che si succedrebbero in un preciso ordine, identico in tutti i bambini e indipendente dalle diverse lingue e culture. Negli ultimi anni si è cercato, al contrario, di evidenziare le differenze piuttosto che le somiglianze, le strategie, i tempi, i percorsi individuali che ciascun bambino può attuare nel processo di acquisizione del linguaggio.

Con l'espressione «acquisizione del linguaggio» si intende il processo di apprendimento e crescita che porta il bambino a comprendere e riprodurre intenzionalmente una lingua naturale, sia essa una lingua materna (L1) o una lingua seconda (L2), e ad interagire verbalmente con altri individui¹. Il linguaggio è una capacità distintiva della specie umana; acquisire il linguaggio è qualcosa che accade in maniera naturale nei primi anni di vita perché gli esseri umani ne sono predisposti.

¹Gian Luigi Beccaria, *Dizionario di linguistica*, Torino, Einaudi, 2004

2.2 Definizione di linguaggio

Definire il linguaggio è molto complesso, in quanto si tratta di un termine inflazionato, al punto che si parla di linguaggio dei media, linguaggio dei giovani, linguaggio del corpo, linguaggio degli animali.

Il linguaggio, nella sua accezione più generale, può essere inteso come un «sistema simbolico di comunicazione», ossia come un sistema in cui l'informazione che passa tra un emittente e un destinatario è codificata in modo simbolico. Ancora più astrattamente e in generale, il linguaggio può essere inteso come la facoltà di comunicare simbolicamente². Naturalmente vi possono essere molti tipi di «linguaggio», a seconda delle possibili combinazioni di emittenti/destinatari, delle condizioni della comunicazione e della varietà delle strutture simboliche impiegate.

Gli studiosi attribuiscono carattere ampiamente specifico alla struttura del linguaggio umano; dunque, le diverse manifestazioni di linguaggio pur essendo sistemi di comunicazione, in quanto trasmettono informazioni da un emittente ad un ricevente o destinatario, sono da differenziarsi dal sistema del linguaggio umano.

Il linguaggio umano è un sistema di simboli che gli uomini usano per comunicare con gli altri. Esso consiste di suoni che possono essere combinati in vari modi per formare le parole, ognuna delle quali è, a sua volta, portatrice di propri, specifici significati. Esistono regole grammaticali che disciplinano il modo in cui le parole possono essere messe insieme per costruire frasi corrette e norme sociali che regolano l'uso appropriato del linguaggio in situazioni differenti. Ciò che sorprende del linguaggio umano è che, sebbene esista soltanto un numero finito di parole e di regole, vi è un numero potenzialmente infinito di modi in cui le parole si possono combinare.

Nel linguaggio umano il simbolo consente di rappresentare a livello astratto la realtà anche in assenza dell'oggetto corrispondente alla realtà stessa. Gli animali sono in grado di arrivare a questo? In merito a ciò, numerosi studiosi hanno cercato di approfondire se le forme di comunicazioni animali sono paragonabili al linguaggio umano.

Una varietà di studi dimostra che gli animali non possiedono la comunicazione simbolica, ma soltanto un sistema di segnali corrispondente ad una serie di manifestazioni comportamentali. Gli animali sono intenzionali e causali, ma non hanno la capacità di rappresentarsi mentalmente gli stati mentali degli altri, né di assumere il punto di vista dell'altro. La comunicazione animale si ferma ai comportamenti manifesti, all'azione, al presente, ma non giunge ad elaborare progetti di vita, ideali, contenuti teorici. Risponde agli stimoli esterni, ma la condotta non è riflessiva, non va oltre il contingente e l'immediato, non ha prospettiva di futuro,

²Il termine «simbolo» va inteso come un tipo particolare di segno. Il concetto è di lunga tradizione nel pensiero occidentale, in quanto risale ad Aristotele (384-322 a.C.), che nel *De interpretatione*, sostiene che «i suoni emessi dalla voce sono i simboli degli stati dell'anima, e le parole scritte sono i simboli delle parole emesse dalla voce».

non ha contenuto simbolico. Per questi motivi si afferma che la comunicazione animale ha per sua natura un carattere imperativo, significa rivolto solo al presente e all'azione, e un valore relazionale.

Infine, la comunicazione animale è esclusivamente non verbale in quanto le vocalizzazioni sono da intendersi come segnali per raggiungere degli scopi e in funzione delle varie situazioni e non come capacità teorica di ricorrere a simboli. In conclusione, i sistemi di comunicazione animale permettono di trasmettere informazioni circa bisogni immediati (pericolo, cibo) e sono relativi all'ambiente circostante. Gli animali comunicano attraverso codici verbali o di altro tipo ma difficilmente siamo disposti ad affermare che queste forme di comunicazione sono paragonabili al linguaggio umano.

Il linguaggio umano è incredibilmente ricco e creativo e permette all'uomo di descrivere fatti e situazioni, informare, mentire, ingannare, persuadere, immaginare situazioni impossibili, esprimere pensieri, slegarsi dal qui e ora e parlare del passato e del futuro; di parlare con chi ha di fronte, ma anche con chi dista mille miglia. Infine, col linguaggio l'uomo può parlare del linguaggio stesso, cioè può attraverso il linguaggio riflettere sul linguaggio (metalinguaggio).

Il ruolo principale negli studi del linguaggio è assunto dalla linguistica, disciplina scientifica e descrittiva (non normativa) che si occupa dello studio della lingua come sistema e dunque di suoni e simboli indipendentemente da chi la usa e dal modo in cui viene usata. Lo scopo conoscitivo (e non pratico) di questa disciplina non è quello di indicare "ciò che si deve dire e non si deve dire" come le grammatiche, ma spiegare ciò che effettivamente si dice. La linguistica include lo studio delle quattro componenti del linguaggio: la fonologia, la semantica, la sintassi e la pragmatica.

2.3 Differenza tra lingua, linguaggio e comunicazione

Spesso la parola «lingua» e la parola «linguaggio» vengono utilizzati come sinonimi nella lingua italiana, ma in realtà non lo sono. Se cerchiamo una definizione di **lingua** nel *Vocabolario Treccani* troviamo "*Sistema di suoni articolati distintivi e significanti (fonemi), di elementi lessicali, cioè parole e locuzioni (lessemi e sintagmi), e di forme grammaticali (morfemi), accettato e usato da una comunità etnica, politica o culturale come mezzo di comunicazione per l'espressione e lo scambio di pensieri e sentimenti, con caratteri tali da costituire un organismo storicamente determinato, con proprie leggi fonetiche, morfologiche e sintattiche*".

Diversa è, invece, la definizione di **linguaggio**: "*[...] la capacità e la facoltà, peculiare degli esseri umani, di comunicare pensieri, esprimere sentimenti, e in genere di informare altri esseri sulla propria realtà interiore o sulla realtà esterna, per mezzo di un sistema di segni vocali o grafici*".

Questa distinzione terminologica esiste solo in alcune lingue: in inglese la parola «language» significa sia «lingua» che «linguaggio» e così accade anche in altre

lingue come il tedesco e il russo. Viceversa, in francese e in italiano, esistono due termini distinti.

La lingua è la forma specifica, storicamente determinata, che il sistema di comunicazione assume nelle varie comunità; in altre parole, è il mezzo di natura verbale con cui si manifesta il linguaggio. «Storicamente determinata» significa che la lingua cambia nel corso della storia in quanto gli eventi storici la condizionano e ne stimolano o ne ostacolano lo sviluppo. La lingua è, quindi, un sistema formale di espressione verbale e scritta che poggia su regole grammaticali, fonetiche, sintattiche, ben precise, di cui si occupa la linguistica e tali regole sono condivise dall'intera popolazione che utilizza quella lingua.

Il linguaggio è, invece, il sistema di comunicazione che ogni essere umano si serve per parlare con altri individui; è una forma di comunicazione tra due o più individui attraverso un complesso determinato di suoni, gesti, simboli e movimenti dotati di significato, che definiscono una lingua comune ad uno specifico ambiente di interazione. Il linguaggio non è, quindi, prerogativa umana: anche le forme di comunicazione animale sono intese come «linguaggio».

Nel modo ordinario di esprimersi si usa spesso impropriamente anche il termine **comunicazione** che stando a quanto riportato nel *Vocabolario Treccani* è “[...] l'azione, il fatto di comunicare, cioè di trasmettere ad altro o ad altri”, “In senso più generale, ogni processo consistente nello scambio di messaggi, attraverso un canale e secondo un codice, tra un sistema (animale, uomo, macchina, ecc.) e un altro della stessa natura o di natura diversa”. La comunicazione è, dunque, il processo che consente di trasmettere informazioni. Comunicare in modo efficace significa sapersi esprimere in ogni situazione con qualunque interlocutore sia a livello verbale che non verbale, in modo chiaro e coerente con il proprio stato d'animo. La comunicazione è uno strumento che va ben oltre le parole, è più ampio perché comprende espressioni verbali, gesti, tono della voce, sguardo, postura.

2.4 Tratti distintivi del linguaggio umano

Come spiegato precedentemente, il linguaggio umano è dotato di caratteristiche proprie, specifiche della specie umana, che sono state individuate per distinguerlo in particolare dalla comunicazione animale e in generale dagli altri sistemi semiotici.

È al linguista americano Charles F. Hockett (1960) che può essere attribuita la paternità dell'enumerazione di tali proprietà da lui denominate *duality, arbitrariness, productivity, interchangeability, specialization, displacement, cultural transmission ecc.*: tali proprietà non ricorrono, nella loro globalità, in nessuna forma di comunicazione diversa da quella umana, anche se possono figurare singolarmente in qualcuna di esse. In tempi successivi, e ad opera dello stesso Hockett e poi di diversi altri studiosi, il numero di tali proprietà è stato variamente riconsiderato

(Hockett in particolare implementò gli originari 13 tratti portandoli a 16); la relativa terminologia è stata in parte rivista e la portata complessiva di tali specificità è stata ripensata.

Alcuni dei tratti distintivi del linguaggio umano caratterizzano solo la modalità orale del linguaggio umano, ma non quella scritta e non si possono estendere alle lingue dei segni, che sono espresse attraverso il canale visivo-gestuale, pure essendo lingue naturali e avendo le stesse caratteristiche del linguaggio orale.

Il linguaggio umano è evanescente: una volta che una frase è stata prodotta non c'è più. Questa proprietà, l'evanescenza, non vale per il linguaggio scritto.

Un'altra proprietà del linguaggio orale è l'intercambiabilità: gli esseri umani possono sia comprendere che produrre il linguaggio, cioè possono assumere il ruolo sia dei parlanti che degli ascoltatori; e in effetti durante uno scambio comunicativo i due ruoli si alternano. Il parlante sente ciò che sente l'ascoltatore, ovvero il parlante sente esattamente ciò che dice nello stesso modo in cui lo sente l'ascoltatore. Ciò non vale per le lingue segniche: chi segna non vede un'immagine speculare di ciò che segna. In altre parole, nelle lingue segniche non c'è il feedback completo.

Il linguaggio umano è un sistema simbolico, in cui le parole e le frasi possono assumere vari significati (semanticità). Ciò non vale per le parole di classe chiusa, cioè articoli e ausiliari. Le parole hanno lo stesso significato, indipendentemente da come sono pronunciate, ad alta o bassa voce, lentamente o velocemente (specializzazione).

Un'ulteriore nota distintiva si riferisce alla discretezza. Nel linguaggio umano, infatti, tutti gli elementi, come ad esempio i suoni, si distinguono in modo ben definito gli uni dagli altri. Suoni diversi, per quanto molto simili, generano parole con significati diversi. La creatività del linguaggio risiede proprio nella possibilità di combinare elementi discreti in modi infiniti.

Un'altra componente che contraddistingue il linguaggio umano è denominata doppia articolazione e indica la possibilità di formare un numero altissimo di segni mediante un numero molto limitato di elementi (i fonemi) che non hanno significato, ma solo la capacità di distinguere significati.

Una delle caratteristiche più squisitamente linguistiche del modo in cui possiamo combinare le parole è la ricorsività. A partire da una frase come a) "Gianni invita Maria", possiamo costruire una seconda frase di cui la prima fa parte b) *Paolo ha detto che* "Gianni invita Maria" ecc. In questo modo, possiamo costruire frasi di lunghezza infinita, semplicemente incassando una frase dentro l'altra. La regola che permette di ottenere queste frasi di lunghezza crescente gode della proprietà ricorsiva, in quanto viene applicata al risultato di una sua applicazione.

L'ultima proprietà del linguaggio che resta da citare è che esso permette di riferirci a concetti e situazioni astratte; in altre parole, il linguaggio consente di staccarci da ciò che è presente e concreto.

Alcune di queste proprietà non sono specifiche del linguaggio, ma sono presenti anche in vari sistemi di comunicazione animale. Altre proprietà sono, inve-

ce, tipiche del linguaggio umano, ma alcune di esse possono essere apprese dagli animali.

Un noto esempio di sistema di comunicazione animale che presenta proprietà in comune con il linguaggio umano è rappresentato dalla «danza delle api», studiato dall'etologo austriaco Karl von Frish (1950, 1967). Le api comunicano fra loro attraverso la danza per fornire indicazioni in merito alla distanza, alla dislocazione e alla qualità del nettare. La danza circolare segnala la presenza di cibo in vicinanza dell'alveare, mentre la danza dell'addome, che ha la forma di otto, indica la direzione e il luogo specifico del cibo quando questo si trova ancora lontano dal luogo di danza. Però, tale danza non ha caratteristiche di discretezza. Per esempio, l'indicazione della distanza avviene attraverso la velocità della danza: quanto più è veloce tanto più è distante la fonte di nettare. La danza delle api non ha dualità di struttura e non è neppure ricorsiva. Questo sistema di comunicazione, accanto ad altri esempi riportati dagli studiosi e presenti in letteratura, pongono in evidenza l'esistenza di una comunicazione animale come sistema organizzato e differenziato e la presenza di competenze comunicative che si dispiegano a più livelli. Tuttavia, non si possono qualificare queste competenze come messaggi veri e propri, poiché, nonostante la capacità di identificare con precisione un oggetto rispetto ad altri presenti nell'ambiente (comunicazione referenziale) e la presenza di uno scopo nell'impiego dei segnali (comunicazione intenzionale), la comunicazione animale non giunge, tanto sul piano della complessità quanto sul piano dell'astrattezza, al livello della comunicazione simbolica.

3

L'ACQUISIZIONE DEL LINGUAGGIO

«La conoscenza che viene acquisita con l'obbligo non fa presa nella mente. Quindi non usate l'obbligo, ma lasciate che la prima educazione sia una sorta di divertimento; questo vi metterà maggiormente in grado di trovare l'inclinazione naturale del bambino»

Platone [La Repubblica]

3.1 Come avviene il processo di acquisizione del linguaggio

Ogni essere umano, all'interno della propria famiglia, è esposto ad almeno una varietà di lingua, utilizzata dai genitori o da altri familiari. Tale varietà di lingua viene definita L1, ovvero lingua materna o lingua prima, essa viene acquisita spontaneamente, attraverso un processo graduale definito acquisizione linguistica, ed è indipendente dall'eventuale istruzione ricevuta o che si riceverà. I bambini sono partecipi, attivi e creativi nel processo di acquisizione della lingua materna. Con il termine L2 o lingua seconda, si intende invece, in linguistica e in glottodidattica, qualsiasi lingua che venga appresa in un secondo momento rispetto alla lingua materna.

Come spiegato nel capitolo precedente, il linguaggio è una capacità distintiva della specie umana: gli animali comunicano, sono intelligenti ma non hanno linguaggio. Acquisire il linguaggio è qualcosa che accade in maniera naturale nel corso dei primi anni di vita perché gli esseri umani ne sono predisposti. Il linguaggio è come un istinto: ne consegue che gli esseri umani non possono esimersi dall'acquisire il linguaggio.

L'acquisizione del linguaggio inizia alla nascita o forse nel grembo materno. Il neonato già a tre giorni di vita è sensibile a vari aspetti del linguaggio; nel corso del primo anno si fa strada nella Babele dei suoni, spezza le stringhe sonore prodotte dai vari parlanti in unità discrete e comincia ad associare alle parole un significato. Questo processo continua durante il secondo anno, e forse non ho mai termine, dal momento che impariamo sempre nuove parole. Alla fine del primo anno di vita,

il bambino inizia a combinare le parole in frasi; poi le frasi diventano sempre più lunghe e complesse. Non è così facile però perché ci sono costruzioni difficili da comprendere, ci sono frasi ambigue localmente o globalmente. Malgrado variazioni nel tipo di informazioni che il bambino può elaborare, nella quantità di input o nel tipo di lingua a cui è esposto (lingua orale o lingua dei segni), tutti i bambini con sviluppo tipico imparano a parlare. Chi si trova nella prima infanzia usa le parole solo per comunicare i propri bisogni primari ma, prima di diventare adulto, può esprimere in modo approfondito idee e concetti astratti. I bambini iniziano a sviluppare il linguaggio nominando e descrivendo gli oggetti; alla fine saranno capaci di esprimere in modo più approfondito idee e concetti astratti e la loro comunicazione non dipenderà dall'ambiente immediato.

3.2 Principali teorie dell'acquisizione del linguaggio

Stabilito che cos'è il linguaggio, quali sono i tratti distintivi che lo caratterizzano e in che modo avviene il processo di acquisizione del linguaggio, in questa parte di ricerca verranno analizzate le prospettive teoriche che cercano di spiegare come esso si sviluppa.

Il dibattito sull'acquisizione del linguaggio ha per lungo tempo ruotato attorno a due concetti, innato e appreso, spesso considerati in opposizione. Secondo gli innatisti, che si ispirano al pensiero di Chomsky, l'acquisizione del linguaggio è resa possibile dall'esistenza di una capacità innata specificatamente predisposta a tale acquisizione. Dall'altra parte i comportamentisti sostenevano la teoria secondo la quale il linguaggio era appreso per imitazione o attraverso il meccanismo del rinforzo casuale di stimolo e risposta. Accanto a questi due principali approcci, numerose sono le teorie che spiegano come avviene il processo di acquisizione del linguaggio, come le posizioni emergentiste e cognitivo-funzionali, il cui denominatore comune è l'ipotesi che esista sì una predisposizione innata all'acquisizione del linguaggio, ma che tale acquisizione sia resa possibile dall'uso di procedure generali non specifiche al linguaggio e che il bambino non sia dotato di una guida innata specifica per acquisire il linguaggio.

3.2.1 La proposta comportamentista

Secondo l'approccio comportamentista, il linguaggio è appreso attraverso il meccanismo del rinforzo casuale dell'associazione tra stimolo e risposta, un meccanismo generale alla base dell'apprendimento di varie abilità, negli esseri umani e negli animali.

Stando a quanto sostiene B. F. Skinner³, i bambini imparano il linguaggio attraverso rinforzi che porterebbero ad una progressiva correzione e strutturazione linguistica: riceverebbero un rinforzo positivo dai genitori quando producono espressioni corrette e un rinforzo negativo quando producono espressioni non corrette. Secondo questa teoria, quindi, il rinforzo positivo aumenterebbe la probabilità che la risposta rinforzata venga ripetuta rendendola una risposta predominante; il rinforzo negativo causerebbe l'inibizione del comportamento. In questo modo grazie all'interazione fra gli stimoli discriminanti e rinforzi il bambino apprende determinati comportamenti (abitudini). Tale teoria riconosce un ruolo passivo dei bambini nel processo di apprendimento linguistico.

Due anni dopo l'uscita del libro di Skinner, nel 1959, appariva la critica di Chomsky che è stata fatale al comportamentismo e che ha stimolato una serie di nuovi studi sul linguaggio. In particolare, Chomsky osservò che Skinner aveva completamente ignorato uno degli aspetti più affascinanti del linguaggio: la creatività, che ci porta a formulare frasi mai sentite prima e la cui espressione non è motivata in alcun senso da uno stimolo. Diversi esempi hanno dimostrato che i bambini producono frasi che non hanno mai sentito, per le quali non possono avere ricevuto alcun rinforzo. È dunque evidente che il meccanismo dello stimolo-risposta non permette di spiegare come avviene l'acquisizione del linguaggio.

3.2.2 La teoria cognitivo-funzionalista

Secondo l'approccio cognitivo-funzionalista, esiste una predisposizione innata per il linguaggio, ma ciò che consente tale acquisizione sono meccanismi generali di cui il bambino è dotato dalla nascita e che vengono usati nell'acquisizione del linguaggio.

Tomasello⁴ (2003) sostiene, infatti, che il linguaggio è acquisito sulla base di meccanismi generali che esistono in altre specie e che svolgono anche funzioni non linguistiche. Tali meccanismi sono di due tipi: lettura delle intenzioni altrui e meccanismi estrattori di configurazioni (pattern). Il bambino, infatti, impara a strutturare gli enunciati non grazie alla comprensione delle regole base della sua lingua ma a partire da item specifici acquisiti nelle interazioni di cui solo lentamente arriverà a generalizzare la struttura. Dunque, secondo questa prospettiva, acquisire il linguaggio comporterebbe dapprima l'acquisizione di espressioni specifiche, ciascuna isolata dall'altra; successivamente e gradualmente, il bambino, sulla base di un'analisi distribuzionale motivata funzionalmente, estrarrebbe delle regolarità, degli schemi che gli consentano di costruire una conoscenza astratta.

³Burrhus Frederic Skinner (1904-1990) è uno psicologo statunitense altamente influente nell'ambito del comportamentismo.

⁴Michael Tomasello (1950) è uno psicologo statunitense, critico verso il concetto di grammatica universale portato avanti da Noam Chomsky.

3.2.3 La teoria emergentista

L'idea fondante delle teorie emergentiste del linguaggio e del suo sviluppo è che le conoscenze sono il risultato dell'interazione fra le strutture attuali del sistema cognitivo e le caratteristiche del mondo fisico, mediante azioni senso-motorie ed operatorie del bambino sul mondo.

Secondo i sostenitori della teoria emergentista, i principi della grammatica, pur non essendo ricavabili dagli stimoli fisici, non sono neppure codificati nel DNA umano. Piuttosto, emergono come la migliore soluzione a problemi specifici posti al genere umano (Bates, Goodman, 1999; Volterra, Bates, 1995): quello di comunicare un vasto repertorio di significati attraverso un sistema cognitivo che presenta delle limitazioni. Secondo questa prospettiva, l'acquisizione del linguaggio è resa possibile da principi cognitivi generali che operano sia nello sviluppo della comunicazione prelinguistica che in quello della comunicazione linguistica. Pertanto, si ritiene che esista continuità tra la prima e la seconda. Inoltre, si ritiene che esista anche continuità tra l'acquisizione del lessico e quella della sintassi nel senso che le abilità grammaticali, la sintassi, emergono dall'acquisizione lessicale. Sulla base di questa prospettiva, si può pensare che i bambini che iniziano a combinare prima, sono quelli che più precocemente hanno sviluppato una sensibilità alla sintassi. La sensibilità per la sintassi fornirebbe loro un meccanismo potente per apprendere nuove parole.

Una prospettiva diversa è stata elaborata da Karmiloff-Smith (1992). Secondo l'autrice la mente dei bambini è costituita da un certo numero di moduli indipendenti e specializzati per specifiche funzioni. Durante lo sviluppo, i moduli si organizzano indipendentemente gli uni dagli altri. Il bambino, secondo Karmiloff-Smith, sarebbe quindi geneticamente predisposto ad assorbire e organizzare le informazioni in strutture modulari, ma è l'ambiente che guida lo sviluppo di tali strutture modulari: sono, infatti, gli input ricevuti dall'esterno che determinano la creazione, l'attivazione e le interazioni di specifici moduli.

3.2.4 La teoria innatista

Secondo la prospettiva innatista, il linguaggio non è un fatto culturale ma una capacità biologicamente determinata; gli esseri umani non sono solo predisposti ad acquisire linguaggio, ma più specificamente dispongono di un qualcosa presente già alla nascita come patrimonio genetico.

L'ipotesi che ispira la teoria innatista è che gli esseri umani affrontino l'acquisizione del linguaggio dotati di una capacità specifica e geneticamente determinata per apprendere il linguaggio che consiste di una grammatica universale (GU).

All'interno di questo approccio, lo studioso per eccellenza è Noam Chomsky, il quale assume l'esistenza di un dispositivo innato per l'acquisizione del linguaggio

(LAD⁵), un programma biologico per imparare a parlare, guidato dalla GU, contenente la descrizione degli aspetti strutturali condivisi da tutte le lingue naturali. In altre parole, il LAD è una struttura cognitiva innata, ipotizzata in modo astratto come funzionante già alla nascita che permette l'acquisizione del linguaggio e lo sviluppo di grammatica e sintassi. All'interno di questo LAD si ipotizza l'esistenza della GU presente in tutti i bambini che si attiverebbe per qualsiasi tipo di lingua. La GU consiste di un insieme di conoscenze astratte e procedure che specificano la particolare forma delle regole linguistiche e quindi la forma possibile o la struttura astratta di tutte le lingue. Guida il processo di acquisizione del linguaggio ed è responsabile del fatto che il linguaggio si acquisisce in tempi e modi identici, indipendentemente dalle variazioni nell'ambiente. La grammatica universale consta di una componente universale che specifica gli aspetti comuni a tutte le lingue e una componente che codifica tutte le variazioni possibili nelle lingue umane.

Per concludere, secondo questa teoria, il processo di acquisizione del linguaggio non consiste nell'imitazione degli adulti, ma è un procedimento attivo di scoperta di regole e di verifica di ipotesi.

⁵LAD: *Language Acquisition Device*, tale meccanismo permette ai bambini di ricavare la struttura sintattica e le regole della loro lingua nativa rapidamente con precisione a partire da un input impoverito.

4

FENOMENO LINGUISTICO DELL'ELLISSI

4.1 Cos'è l'ellissi

Il termine ellissi, dal greco antico ἔλλειψις, *èlleipsis*, “mancamento”, “omissione”, derivato del verbo ἔλλείπω, *èllèipō*, “io ometto”, significa letteralmente mancanza. Si tratta di una figura retorica che consiste nell'omissione, all'interno di una frase, di una o più parole che la costruzione grammaticale richiederebbe e che sia possibile sottintendere.

Secondo quanto affermato da Charlotte Lindenbergh, Angeliek van Hout, and Bart Hollebrandse (University of Groningen) nell'articolo *Extending ellipsis research: The acquisition of sluicing in Dutch* l'ellissi è il fenomeno linguistico in cui manca la struttura linguistica (suono o scrittura), ma l'interpretazione rimane comunque.

Nei dizionari moderni, come ad esempio nel *Vocabolario Treccani*, l'ellissi viene definita come “l'omissione di qualche elemento che resta sottinteso in una frase e risulta ricavabile dal contesto”.

In vari quadri teorici l'ellissi viene analizzata come un tipo speciale di anafora, perché il significato della parte muta di una frase ellittica deve essere recuperato dal contesto del discorso. Sono stati identificati diversi tipi di ellissi, a seconda di quanta struttura viene tralasciata. Il luogo in cui manca la struttura è spesso chiamato il sito di ellissi, indicato da '___' nelle frasi di esempio seguenti.

(i) *NP Ellipsis + non-elliptical counterpart*

- Sam is waving with two hands, and Anna waves with one ___.
- Sam is waving with two hands, and Anna waves with one hand.

(i) *VP Ellipsis + non-elliptical counterpart*

- Mark and Sam play the guitar, but Anna doesn't ___.
- Mark and Sam play the guitar, but Anna doesn't play the guitar.

(i) *Sluicing + non-elliptical counterpart*

- Someone is making noise, but I don't know who ____.
- Someone is making noise, but I don't know who is making noise.

La ricerca sull'acquisizione dell'ellissi si è concentrata principalmente sull'ellissi verbale (VPE – *verb phrase ellipsis*) e nominale (NPE – *noun phrase ellipsis*), e solo uno studio ha esaminato l'acquisizione dello sluicing.

Un fenomeno parzialmente distinto dall'ellissi è l'omissione del soggetto che, pur rappresentando un caso di ellissi, non si tratta di una semplice cancellazione: in lingue come l'italiano, in cui il verbo accorda in persona, numero ed eventualmente genere con il soggetto, il soggetto non espresso è marcato dalla flessione di persona del verbo, ovvero esiste un mezzo grammaticale di segnalazione del referente testuale.

4.2 Rassegna dei risultati di studi precedenti

In letteratura sono state proposte diverse analisi teoriche sull'ellissi di varie lingue. Le principali domande nella ricerca teorica sull'ellissi volte a capire l'utilizzo di questo fenomeno linguistico da parte dei bambini sono se sia presente o meno una struttura sintattica nel sito dell'ellissi e in quali contesti l'ellissi sia autorizzata.

Dall'analisi dei risultati di studi precedenti sull'ellissi risulta evidente che i bambini piccoli sono già abbastanza in grado di trattare con questo fenomeno linguistico e che il discorso diretto dei bambini contiene molte strutture ellittiche. Tuttavia, la produzione di una struttura ellittica non implica la comprensione o il pieno controllo della struttura stessa, infatti come mostrano i dati del Corpus è evidente che l'utilizzo di ellissi nel discorso diretto dei bambini può anche portare ad un fallimento nella comunicazione e di conseguenza comprensione tra adulto-bambino.

Mentre l'ellissi è frequente nell'input linguistico dei bambini e produce alcune semplici forme di ellissi già dall'età di due anni, padroneggiare tutte le strutture ellittiche ed essere in grado di ricostruire le informazioni linguistiche precedenti dal discorso potrebbe non essere così semplice per i bambini piccoli. A questo proposito numerosi studiosi hanno cercato di fornire diverse analisi teoriche riguardanti l'acquisizione dell'ellissi da parte del bambino nel dialogo con i *caregivers*.

Nei paragrafi successivi segue l'analisi di due articoli contenenti i risultati di studi riguardanti l'ellissi spontanea nei bambini.

4.2.1 *The acquisition of nominal ellipsis in Greek*

La *noun ellipsis*, chiamata anche *nominal ellipsis*, nell'articolo *Acquisition of noun ellipsis in French and in Dutch – consequences for linguistic theory* di Petra Sleeman e Aafke Hulk (*ACLCL, University of Amsterdam*) viene definita come "l'omis-

sione del sostantivo in un contesto in cui l'interpretazione del sostantivo eliso può essere recuperata dal contesto linguistico o situazionale". Si tratta dunque di una struttura sintattica in cui la testa nominale di una frase nominale è stata omessa.

Secondo quanto riportato da Dimitris Ntelitheos e Eleni Christodoulou, un certo numero di approcci ha associato l'ellissi nominale ad un accordo morfologico "ricco" nel dominio nominale (Lobeck 1995, Kester 1996). Se la morfologia ricca permette l'ellissi nominale, allora ci si aspetterebbe che l'ellissi nominale appaia nelle espressioni dei bambini non prima dello stadio in cui l'accordo morfologico nel dominio nominale è pienamente sviluppato. Questa previsione è stata testata sulla base dei dati dell'acquisizione del greco, essendo una lingua con un sistema di accordo molto ricco nel dominio nominale.

I dati utilizzati nello studio sull'accordo nominale nel discorso dei bambini greci consistono in una parte dello Stephany Corpus, un corpo longitudinale di quattro bambini greci monolingui, disponibile nel database CHILDES (MacWhinney & Snow 1985, Stephany 1997). Dall'analisi dei dati è stato dimostrato che il discorso dei bambini greci in questa fase (1;9-1;11) è composto da frasi nominali che contengono teste nominali e modificatori che concordano solo in casi non marcati. I bambini in questa fase producono frasi nominali piuttosto complesse con un numero di modificatori prenominali. Gli ambienti in cui l'ellissi nominale appare nel discorso dei bambini sono diversi come gli ambienti nel discorso degli adulti: l'ellissi nominale appare con aggettivi (*'I will buy a new one'*), quantificatori (*'I have many'*) e possessivi (*'I will buy my own one'*), per cui è presente in modo produttivo molto presto nel discorso dei bambini greci.

Sembra, quindi, che la presente ricerca falsifichi l'ipotesi secondo cui l'accordo morfologico "ricco" nel dominio nominale sia un prerequisito o un meccanismo che abilita l'ellissi nominale in quanto è stato concluso che i bambini non hanno ancora appreso il paradigma completo dell'accordo nominale nel dominio nominale, o che sono nelle prime fasi di apprendimento di un tale paradigma.

Un ulteriore argomento a sostegno di questa conclusione deriva dal fatto che i bambini producono costruzioni ellittiche elidendo un sostantivo che ha caratteristiche incompatibili con le caratteristiche del modificatore. In altre parole, i bambini sembrano produrre ellissi nominali con un accordo errato tra il sostantivo eliso e il restante modificatore.

In conclusione, queste analisi rafforzano l'ipotesi che la "ricca" morfologia dell'accordo non è il meccanismo che abilita l'ellissi nominale.

4.2.2 *Acquisition of noun ellipsis in French and in Dutch - consequences for linguistic theory*

In questo articolo di Petra Sleeman e Aafke Hulk (*ACLIC, University of Amsterdam*) vengono presentati e discussi i risultati di uno studio esplorativo sull'acquisizione precoce dei sostantivi elisi con e senza il pronome quantitativo da parte di

bambini monolingui francesi e olandesi (fascia di età compresa tra 1.8 - 3.6) sulla base di produzioni spontanee.

È ben noto che nelle prime fasi dell'acquisizione del linguaggio i bambini non producono ancora DP completi: omettono determinanti e aggettivi, producendo solo il sostantivo. Tuttavia, è meno noto che i bambini producono anche DP contenenti un aggettivo o un quantificatore, ma nessun sostantivo. Sulla base del contesto in cui questi DP vengono prodotti, essi sembrano essere veri e propri casi di ellissi nominale e non DP incompleti prodotti erroneamente o nominalizzazioni.

Secondo la letteratura gli elementi che possono svolgere un ruolo nella legittimazione di *noun ellipsis* sono la partitività, la flessione aggettivale e/o il pronome quantitativo. Soprattutto per i primi due elementi esistono affermazioni contraddittorie in letteratura rispetto alla loro rilevanza nell'acquisizione di L1. Per testare la rilevanza di questi tre elementi, Petra Sleeman e Aafke Hulk hanno studiato di dati di produzione spontanea di diversi bambini monolingui L1 francesi e olandesi.

È stato dimostrato che, proprio come nel francese adulto e nel bambino L1 francese, nel bambino L1 l'ellissi nominale olandese sembra essere autorizzata da partitività. I dati hanno dimostrato che nel bambino L1 l'ellissi olandese può essere prodotta anche in assenza di un determinante e di (corretta) inflessione aggettivale. Pertanto, le teorie sull'ellissi nominale che rivendicano un ruolo cruciale per i determinanti e/o la flessione dovrebbero essere riesaminate. Inoltre, è stato dimostrato che l'acquisizione del ruolo del pronome quantitativo nel meccanismo di legittimazione delle ellissi nominali procede sostanzialmente allo stesso modo in francese e in olandese, supportando per entrambe le lingue un'analisi del movimento di una variante lessicale di un sostantivo vuoto. Si tratta tuttavia di un lavoro in corso e quindi questi risultati possono solo suggerire alcune risposte.

5

ANALISI DEI DATI

Quest'ultima parte di tesi è dedicata ad un capitolo sperimentale di analisi dei dati in cui verranno analizzate e discusse le produzioni input contenenti ellissi a cui i bambini sono effettivamente esposti durante il dialogo con i caregivers. In seguito, verranno analizzate le produzioni output prodotte dai bambini. Per concludere i dati verranno confrontati per capire se correlano per qualità, cioè per tipologia, e per quantità. Per l'analisi dei dati è stato utilizzato un corpus, ovvero una collezione di testi selezionati e organizzati per facilitare le analisi linguistiche; in questo caso il Childes Italian Tonelli Corpus.

5.1 La banca dati CHILDES

I primi studi svolti sull'acquisizione del linguaggio si basavano principalmente sull'osservazione, da parte dei ricercatori, di un ristretto numero di bambini (spesso i loro figli) di cui monitoravano e trascrivevano ogni frase o parola pronunciata durante le interazioni familiari. Questo studio diaristico, per la sua complessità, poteva concentrarsi unicamente su un numero circoscritto di bambini. Lo sviluppo tecnologico ha poi ovviato a questo limite grazie anche alla creazione di un data base in cui ogni ricercatore può mettere le proprie trascrizioni a disposizione della comunità scientifica, il Child Language Data Exchange System (CHILDES). Tale banca dati è in continua evoluzione e contiene produzioni spontanee di bambini che parlano lingue molto diverse: italiano, cinese, tedesco, inglese, francese, spagnolo, ecc.

I dati all'interno del sistema CHILDES sono organizzati in file e ogni file contiene la trascrizione di ciò che il bambino e uno o più adulti hanno prodotto durante una sessione di gioco. I file hanno la struttura che inizia con l'istruzione di apertura (@BEGIN), seguita da varie informazioni circa i partecipanti, l'età del bambino e la situazione di produzione linguistica. Seguono poi le produzioni dei bambini (CHI) e degli adulti. Il file termina con l'istruzione @END.

5.2 Le fasi dell'analisi dei dati sulla base di CHILDES Italian Tonelli Corpus

L'analisi dei dati è un processo mediante il quale si ricavano informazioni da dati grezzi che vengono estratti, trasformati e analizzati. Per lo svolgimento di questo processo effettuato sulla base di un corpus sono state eseguite diverse fasi di cui si riporta la descrizione.

Il corpus da cui sono stati raccolti i dati è il Childes Italian Tonelli Corpus, strutturato in tre sottocorpora corrispondenti a tre bambini Elisa, Gregorio e Marco di età compresa tra un anno e cinque mesi e due anni e cinque mesi. Ogni sottocorpus contiene la trascrizione di ciò che il bambino e il caregiver, in genere la madre ma nel caso di Gregorio anche la zia, hanno prodotto durante sessioni di gioco.

La prima fase della ricerca è stata l'acquisizione dei dati: dopo aver individuato tutte le produzioni contenenti ellissi pronunciate dai caregivers durante la conversazione con il bambino, i dati sono stati riuniti all'interno di un file in Excel.

Successivamente, è stato analizzato quello che viene tecnicamente chiamato *remnant*, che in italiano corrisponde a "residuo dell'operazione di ellissi". Il *remnant* è quindi la porzione residuale lessicalizzata che rimane dopo l'operazione di ellissi. Durante questa fase di elaborazione dei dati, oltre a trascrivere nel file in Excel il tipo di *remnant*, il tipo di determinante e il token di determinante, è stata verificata la presenza o assenza di modificatori aggettivali o altri modificatori. Guardando alla raccolta dati, sono state tratte delle generalizzazioni e osservazioni confermate poi dall'analisi riassuntiva effettuata attraverso tabelle pivot, uno strumento che ha permesso di raggruppare e visualizzare i dati in modo più chiaro e sintetico.

A seguito dell'analisi delle tipologie di ellissi pronunciate dai caregivers, la domanda che è sorta è se la qualità e la quantità di input del caregiver correlano per qualità e per quantità con l'output del bambino. Per confrontare quello che è l'output del bambino rispetto all'input ricevuto, sono stati inseriti all'interno di un unico file in Excel tutti i *remnant* input pronunciati dai caregivers e tutti i *remnant* output pronunciati dai bambini. Questo lavoro ha permesso di compiere un'analisi più dettagliata per singolo bambino, analizzando separatamente bambino input con bambino output.

Infine, la ricerca si raffina ancora di più andando ad esaminare gli input e gli output di ogni bambino in base all'età.

5.3 Descrizione per qualità e quantità delle produzioni input e output

Tutte le produzioni contenenti ellissi pronunciate dai caregivers di Elisa, Gregorio e Marco sono state inserite all'interno di un file in Excel. In totale le produzioni

input sono 192 suddivise in 9 per Elisa, 5 per Gregorio e 178 per Marco. Facendo la raccolta dati sono emerse delle generalizzazioni e osservazioni che sono state confermate poi dall'analisi riassuntiva effettuata attraverso l'uso di tabelle pivot.

La prima generalizzazione riguarda la tipologia di *remnant* prodotto dai caregivers. Vi sono due tipologie fondamentali: pronomi dimostrativo distale (e.g.⁶ MOT: *poi manca quello rosso*) e pronomi numerale (e.g. MOT: *uno giallo, bravo*). Entrambi i determinanti in alcuni casi vengono pronunciati singolarmente; in altri casi sono accompagnati da aggettivo oppure da altri modificatori. Ci sono rari casi in cui il *remnant* dimostrativo distale appare più complesso e modificato da aggettivo e PP⁷ (e.g. MOT: *ah quello bianco con le macchie nere*).

La Tabella 5.1 mostra il conteggio di quante volte è stata realizzata ellissi divisa per il caregiver di Elisa, Gregorio e Marco e divisa a sua volta per la tipologia di determinante che è stato prodotto. Come si evince dalla tabella, la grande totalità delle produzioni si articolano in dimostrativo distale (84 casi 43,75%) e numerale (46 casi 23,96%), andando così a confermare quanto detto precedentemente. Bisogna però notare che questi tipi di determinante sono presenti in grande quantità nelle produzioni del caregiver di Marco, mentre sono assenti in Elisa e c'è solo un caso di dimostrativo distale in Gregorio. Questo è dovuto al fatto che Marco contiene più produzioni e quindi la quantità di tipo di determinante è maggiore rispetto agli altri bambini.

Conteggio di Tipo_Determinante Etichette di colonne		distale_dimo	indefinit	prossimale_		universale	universale+distale	Totale	
Etichette di riga	definito	strativo	o	numerale	dimostrativo	quantificatore	+definito	_dimostrativo	complessivo
Elisa				3		5	1		9
Gregorio		1	1	1			1	1	5
Marco		9	83	13	46	21	5		178
Totale complessivo		10	84	17	46	26	7	1	192
Totale complessivo in %		5,21%	43,75%	8,85%	23,96%	13,54%	3,65%	0,52%	100,00%

Tabella 5.1: Conteggio di Tipo_Determinante.

Guardando i dati della Tabella 5.1 interessante è la porzione del quantificatore (e.g. MOT: *ne abbiamo tanti*) di cui però ci sono solo 7 casi in totale, così come quella del dimostrativo prossimale (e.g. MOT: *no, questi sono piccoli*) di cui ci sono 26 ricorrenze. Come nel caso del distale dimostrativo, anche il prossimale dimostrativo viene utilizzato senza modificatore, con l'aggettivo come modificatore e con il PP come modificatore (e.g. MOT: *questo della carica dei Centouno*).

Altri determinanti usati dai caregivers sono il definito (e.g. MOT: *dai, mettiamo la blu e la gialla la dentro*) di cui ci sono 10 ricorrenze e l'indefinito (e.g. MOT: *è lo stesso orsetto (.) o è un altro?*) di cui ci sono 17 casi.

⁶E.g. è l'abbreviazione di *exempli gratia*, ovvero *example given* in inglese e "per esempio" in italiano.

⁷Con PP si intendono anche le relative clauses, cioè tutti quei modificatori non aggettivali.

Infine, come si può notare dalla Tabella 5.1, meno ricorrente è l'associazione di universale+definito (i.e.⁸ ANT: *un altro bello (.) qua ci sono tutti gli altri, che belli.*) presente in Gregorio, e l'unione universale+dimostrativo distale (i.e. MOT: *bene mi par(e) che le abbiamo tutte quelle del treno.*) presente in Marco.

La ricerca sulla qualità e quantità di input continua e si raffina andando ad analizzare in modo più dettagliato quali quantificatori e numerali vengono utilizzati e con quale frequenza. A questo scopo è stata creata una tabella pivot per studiare la qualità del token del determinante per solo un subset, cioè un sottoinsieme, di tipi di determinante: la Tabella 5.2 si riferisce all'analisi del quantificatore e la Tabella 5.3 all'analisi del numerale.

Per quanto riguarda l'analisi del quantificatore, come è emerso dalla tabella precedente, sono poche le produzioni che lo contengono (solamente 7 casi su 192 totali). Non considerando la questione del genere, sostanzialmente viene utilizzato un solo quantificatore: *tanto*.

Tipo_Determinante	quantificatore		
Conteggio di Token_Determinante		Etichette di colonna	
Etichette di riga	tanto	Totale complessivo	
Elisa	1	1	
Gregorio	1	1	
Marco	5	5	
Totale complessivo	7	7	

Tabella 5.2: Conteggio di Token_Determinante con filtro su quantificatore.

Tipo_Determinante	numerale				
Conteggio di Token_Determinante		Etichette di colonna			
Etichette di riga	cinque	due	tre	uno	Totale complessivo
Marco	1	5	3	37	46
Totale complessivo	1	5	3	37	46

Tabella 5.3: Conteggio di Token_Determinante con filtro su numerale.

In quanto all'analisi del numerale, come si può notare dalla Tabella 5.3, il numerale ha una frequenza ridotta e concentrata sulle produzioni del caregiver di Marco, con uno sbilanciamento sul numerale *uno* (37 casi su 46 totali).

Le analisi dei dati compiute sulle produzioni input sono state eseguite anche per quanto riguarda le trascrizioni contenenti ellissi pronunciate dai bambini durante

⁸I.e. è l'abbreviazione di *id est*, ovvero *that is* in inglese e "cioè" in italiano.

le conversazioni con i caregivers. La prima differenza che si nota dopo aver inserito tutte le produzioni output all'interno di un file in Excel è la diversa quantità di ellissi: mentre le produzioni input dei caregivers erano in totale 192, le produzioni output dei bambini sono 576. Un aspetto che può giustificare questo è la quantità maggiore di numero di righe dei bambini rispetto ai genitori all'interno del Corpus. Nella maggior parte dei casi è il genitore che fornisce l'antecedente linguistico al bambino, ma è il bambino che, soprattutto con l'aumentare dell'età, parla egocentricamente da solo e produce ellissi sulla base del suo stesso discorso.

Le produzioni output sono così suddivise: 62 per Elisa, 22 per Gregorio e 492 per Marco.

La prima analisi che è stata effettuata riguarda il tipo di *remnant*. Elisa ha tre grandi tipologie di *remnant*: definito modificato da aggettivo (e.g. CHI: *e il mio dov'è?*), numerale (e.g. CHI: *allora uno era qui*) e quantificatore (e.g. CHI: *tutti in casa*). Nelle produzioni di Elisa è presente con una buona percentuale anche il dimostrativo distale (e.g. CHI: *quello lì*). Gregorio, a differenza di Elisa, presenta tre strategie: indefinito modificato con aggettivo (e.g. CHI: *c'è un altro (.) un altro bello*), numerale e quantificatore. In Gregorio è assente, dunque, l'uso del definito modificato con aggettivo. Marco, similmente a Elisa e Gregorio, presenta come strategia il numerale e di conseguenza si può affermare che il numerale è la tipologia di *remnant* condivisa da tutti e tre i bambini. Nelle produzioni di Marco sono presenti con una buona percentuale il quantificatore, anche questa è una strategia condivisa da tutti e tre i bambini, e l'aggettivo (e.g. CHI: *oh verde*), assente in Gregorio e poco presente in Elisa. Quindi, nella maggior parte dei casi, ad eccezione di Gregorio dove mancano delle produzioni, le tre grandi categorie più rappresentate sono: indefinito modificato con aggettivo, numerale e quantificatore. Ci sono poi variazioni tra bambini sia nella qualità, ovvero nella tipologia di *remnant*, sia nella quantità; variazioni che riguardano in particolare Marco che, essendo quello che contiene più produzioni, ha anche un range di qualità maggiore rispetto agli altri.

Come nel caso dell'input, anche per le produzioni di output interessante è analizzare quali quantificatori e numerali vengono utilizzati e con quale frequenza. La Tabella 5.4 si riferisce all'analisi del quantificatore e la Tabella 5.5 all'analisi del numerale.

Per quanto riguarda l'analisi del quantificatore, sono poche le produzioni contenenti quantificatore. Se non viene considerata la questione del genere, sostanzialmente sono cinque i quantificatori: *altro*, *poco*, *quale*, *quanto*, *tanto*. Questi quantificatori hanno una frequenza ridotta, sono presenti in Elisa e Marco, mentre in Gregorio l'uso del quantificatore è assente. Se si guarda alle produzioni più consistenti di Marco, il tutto è molto sbilanciato sul quantificatore *tanto* (nelle varie forme ma sul lessema *tanto*).

Per quanto riguarda l'analisi del numerale, frequente è l'uso di *uno*: 6 casi su 15 totali in Elisa, 3 su 5 totali in Gregorio e 44 su 142 totali in Marco. Similmente in

Marco c'è una grande produzione anche di *due* e *tre*, poi *quattro* e più di progredisce nella scala numerica più si diminuisce la frequenza. Questo aspetto riguardante la tipologia di numero e di ordine della scala numerica è interessante: l'uso di numeri molto bassi è comprensibile dal momento che l'analisi interessa l'output del bambino in una conversazione con il genitore durante una sessione di gioco. I contesti in cui vengono elicitati tipologie di numeri molto alti (e.g. *cento*, *sessanta*) sono ristretti in termini di topic del discorso e di conseguenza sono anche quelli meno frequenti nella produzione del parlante.

Tipo_Determinante		quantificatore									
Conteggio di Token_Determinante		Etichette di colonna									
Etichette di riga		altri									
		poche	pochi	quale	quante	quanti	tante	tanti	tantissimi	Totale complessivo	
Elisa						1		1	1		3
Marco		2	2	1	4	2	1	23	21		56
Totale complessivo		2	2	1	4	2	2	23	22	1	59

Tabella 5.4: Conteggio di Token_Determinante con filtro su quantificatore.

Tipo_Determinante		numerale											
Conteggio di Token_Determinante		Etichette di colonna											
Etichette di riga		cinque											
		dieci	due	nove	otto	quattordici	quattro	sedici	sei	tre	undici	uno	Totale complessivo
Elisa			3	1	2		1		2			6	15
Gregorio				2								3	5
Marco		9	2	43	2	3		12	3	23	1	44	142
Totale complessivo		9	2	48	3	5	1	12	2	3	23	1	53

Tabella 5.5: Conteggio di Token_Determinante con filtro su numerale.

Un'altra analisi che è stata condotta sulla base delle produzioni output riguarda la tipologia di modificatore. Interessante è capire quale modificatore, aggettivo o PP, viene utilizzato con più frequenza dai bambini. Come si può notare dalla Tabella 5.6 nella maggior parte dei casi non c'è modificatore (e.g. CHI: *le voglio tutte*). Nel 40,28% dei casi il *remnant* è modificato da un aggettivo (e.g. CHI: *questa è la mia*) e nel 6,25% da una preposizione (e.g. CHI: *anch' io c(i) ho questa in camera*).

Infine, la ricerca si raffina ancor di più andando ad analizzare quale tipologia di aggettivi e PP viene utilizzata con più frequenza. Dall'analisi dei dati si evince che vi è una grande maggioranza di *altro* e un uso frequente di aggettivi di colore. Altri aggettivi utilizzati sono i possessivi presenti solo per due partecipanti Elisa e Marco e vi è in Marco una produzione di aggettivo di forma (i.e. CHI: *abbiamo di quelli tondi?*). Per quanto riguarda le PP, ci sono PP funzionali introdotte dalla preposizione con (e.g. CHI: *quella con la marmellata*), PP con la preposizione di

Conteggio di NP_MODIFIER_TYPE	Etichette di colonna							
Etichette di riga	0	ADJ	ADV	NUM	PREP	RC	Totale complessivo	
Elisa	41,94%	37,10%	9,68%	1,61%	6,45%	3,23%	100,00%	
Gregorio	63,64%	31,82%	0,00%	4,55%	0,00%	0,00%	100,00%	
Marco	47,56%	41,06%	3,25%	1,02%	6,50%	0,61%	100,00%	
Totale complessivo	47,57%	40,28%	3,82%	1,22%	6,25%	0,87%	100,00%	

Tabella 5.6: Conteggio di Tipo_Modificatore.

(e.g. CHI: *quella della nonna*) e PP con la preposizione per (e.g. CHI: *uno per terra*). Sono poi presenti preposizioni complesse spaziali (i.e. *dentro, dentro al recinto, dietro*) e preposizioni locative (i.e. *in camera, sul ghiaccio, nel recinto*).

5.4 Confronto dei dati input e output

L'ultima parte di questo lavoro di analisi dei dati è stata dedicata al confronto tra produzioni input pronunciate dai caregivers e produzioni output pronunciate dai bambini per esaminare se correlano per qualità, cioè per tipologia, e per quantità. La ricerca è divisa in due parti: nella prima, dedicata ad un'analisi che riguarda ogni singolo bambino, verranno confrontati bambino input e bambino output; nella seconda parte la ricerca si raffina ancora di più andando ad esaminare gli input e gli output di ogni bambino in base all'età.

La prima indagine riguarda il tipo di determinante utilizzato nelle produzioni input e output. Un dato che si evidenzia è la maggior varietà di tipo di determinante utilizzato nelle produzioni output rispetto alle produzioni input, soprattutto nei confronti dei dati di Elisa e Marco. In Elisa input i determinanti utilizzati dal caregiver sono l'indefinito e il quantificatore. Il dimostrativo prossimale, nonostante presenti un'alta percentuale in Elisa input, non viene preso in considerazione in quanto troppo ambiguo per dirimere tra un'interpretazione ostensiva e un'interpretazione anaforica. Diverso è, invece, il conteggio in Elisa output dove i determinanti utilizzati dal bambino sono principalmente numerale, definito e dimostrativo distale assenti nelle produzioni input, e in sempre minor quantità vengono anche utilizzati i determinanti universale, indefinito, quantificatore, dimostrativo prossimale e universale+numerale. Il confronto tra Gregorio input e Gregorio output è inverso: la varietà di tipo di determinante in Gregorio input è maggiore e presenta una percentuale uguale nell'uso di cinque determinanti, tra cui l'indefinito, unico determinante in comune con le produzioni di Gregorio output. I determinanti universale e numerale utilizzati in grande quantità dal bambino durante la conversazione con il caregiver sono totalmente assenti nelle produzioni input. L'ultimo confronto sul tipo di determinante riguarda Marco input e Marco output: anche in questo caso, come in Elisa, la varietà di determinanti utilizzati dal bambino è maggiore rispetto a quelli pronunciati dal caregiver. Il confronto tra le produzioni

di Marco ha evidenziato il maggior utilizzo di numerale e dimostrativo distale sia in input sia in output. In Marco output, e anche in due casi di Elisa output, frequente è l'uso di aggettivo privo di determinante (e.g. CHI: *mi piace gialla mam(m)a*).

Il confronto dei dati input e output continua andando ad analizzare in modo più dettagliato quali quantificatori e numerali vengono utilizzati e con quale frequenza. Per quanto riguarda il confronto tra caregiver e bambino, in Elisa input vi è solamente una produzione che contiene il quantificatore *tanto*, mentre in output se non viene considerata la questione del genere, i quantificatori sono due: *quanto*, *tanto*. In output maggiore è il numero delle produzioni che contiene il numerale (15 casi su 18) rispetto al quantificatore. Diverso è il confronto tra Gregorio input e Gregorio output: mentre il quantificatore è presente in una produzione input e assente in output, il numerale viene utilizzato in cinque produzioni output ed è assente in input. Più diversificata è la situazione in Marco dove sia da parte del caregiver sia del bambino vengono utilizzati quantificatori e numerali differenti. Per quanto riguarda l'analisi del quantificatore, non considerando la questione del genere, in Marco input viene utilizzato solo il quantificatore *tanto*, mentre in Marco output sostanzialmente sono cinque i quantificatori: *altro*, *poco*, *quale*, *quanto*, *tanto*. In quanto all'analisi del numerale, sia in input sia in output frequente è l'uso del numerale *uno*; in output c'è una grande produzione anche di *due* e *tre*, poi quattro e più di progredisce nella scala numerica più si diminuisce la frequenza (vedi Tabella 5.5).

Un'ultima analisi effettuata prendendo in esame ogni singolo bambino e confrontando dati input con dati output riguarda la tipologia di aggettivi e PP utilizzati nelle produzioni durante la conversazione. Dall'analisi dei dati condotta attraverso l'uso di tabelle pivot è emerso che in Elisa input l'uso di aggettivi è limitato e sbilanciato su *altro* e quello di PP è totalmente assente. In Elisa output, invece, per quanto riguarda la presenza di aggettivi vi è una grande maggioranza dei possessivi, mentre per quanto riguarda le PP, ci sono PP funzionali introdotte dalle preposizioni *con*, *di* e un caso di preposizione complessa spaziale (i.e. *in camera*). Più semplice è il confronto tra Gregorio input e Gregorio output in quanto il conteggio della presenza di PP è nullo, ad eccezione di un caso di produzione del caregiver. Mentre per quanto riguarda la presenza di aggettivi, il conteggio è sbilanciato, come nel caso di Elisa, su *altro*. Infine, ampio ed interessante è il confronto tra Marco input e Marco output per la varietà di aggettivi e PP utilizzati sia dal caregiver sia dal bambino. In entrambi i casi vi è una grande maggioranza di *altro* e un uso frequente di aggettivi di colore. Per quanto riguarda le PP, sia in Marco input sia in Marco output ci sono PP funzionali introdotte dalle preposizioni *con* e *di*. Solo in Marco output vi sono casi di PP introdotte dalla preposizione *per*, preposizioni complesse spaziali (i.e. *dentro*, *dentro al recinto*, *dietro*) e preposizioni locative (i.e. *sul ghiaccio*, *nel recinto*).

La seconda parte di questo confronto dei dati esamina gli input e gli output

di ogni bambino in base all'età. Interessante è l'analisi sul tipo di determinante utilizzato nelle produzioni input e output suddivise per fasce d'età del bambino. Dal confronto tra Elisa input ed Elisa output con focus sull'età 1;10 - 1;11 - 2;01⁹ si evince che i tipi di determinante in output variano per qualità, cioè per tipologia, e per quantità rispetto ai corrispettivi in input. Come detto precedentemente, un aspetto che dev'essere tenuto in considerazione e che può giustificare questo squilibrio tra input e output è la quantità maggiore di numero di righe di produzione dei bambini rispetto a quelle dei genitori all'interno del Corpus. Una differenza che si nota nel confronto tra Elisa output con focus sull'età 1;10 - 1;11 ed Elisa output prendendo in considerazione anche l'età 2;01 è la maggior quantità dei determinanti dimostrativo distale e numerale con l'aumentare dell'età. Interessante è il confronto tra Gregorio input e Gregorio output con focus sull'età 1;08 - 1;09 - 1;10 - 2;00 in quanto dall'analisi dei dati si evidenzia che da 2;00 maggiore è la qualità e quantità di determinanti presenti sia nelle produzioni del caregiver sia in quelle del bambino. In Gregorio input con focus sull'età 1;08 - 1;09 - 1;10 vi è un unico tipo di determinante, il dimostrativo distale; mentre in Gregorio output le tipologie di determinanti sono indefinito, universale e numerale. Considerando anche gli input e output relativi all'età 2;00, il conteggio di tipo di determinante aumenta per qualità e quantità, soprattutto nelle produzioni input del caregiver dove da un unico determinante si passa all'uso di cinque determinanti diversi. Ultimo confronto riguarda Marco input e Marco output: considerando la sezione dei dati in base all'età da 1;05 a 1;11, il determinante presente con maggior frequenza è il numerale, seguito in input dal dimostrativo distale e in output da quantificatore e indefinito. Con l'aumentare dell'età e quindi considerando nell'insieme anche i dati relativi all'età da 2;00 a 2;05, la frequenza del dimostrativo distale aumenta sia nelle produzioni input del caregiver sia nelle produzioni output del bambino. Dall'analisi dei dati si può dunque dedurre che l'incremento in qualità e quantità del tipo di determinante è correlato con l'aumentare dell'età del bambino.

In conclusione, per rispondere alla domanda se qualità e quantità di input del caregiver determina qualità e quantità di output del bambino, si può affermare che le produzioni input contenenti ellissi a cui i bambini sono esposti durante il dialogo con i caregivers influenzano parzialmente le produzioni output. I bambini, soprattutto con l'aumentare dell'età, parlano egocentricamente da soli e dall'analisi dei dati è stato constatato che producono ellissi sulla base del loro stesso discorso. Per questo motivo si può sostenere l'ipotesi che i bambini fin dalla tenera età iniziano a produrre determinate strutture sintattiche non solo per imitazione del caregiver.

⁹Con questa dicitura si identifica l'età del bambino, per esempio 1;10 fa riferimento ad un'età di un anno e dieci mesi.

6

CONCLUSION

This elaborate was written to offer a theoretical review on language and its acquisition process and a practical one on the use of ellipsis by parents and children during their dialogues. The first part of the elaborate consisted of three chapters dedicated to theoretical notions on language acquisition, its principal theories and the ellipsis' linguistic phenomenon; the second one, composed by one chapter, was entirely reserved to the analysis of the data gathered from CHILDES Italian Tonelli Corpus. In this experimental chapter of data analysis, the input productions containing ellipsis to which children are exposed during dialogue with caregivers were analyzed and discussed. Subsequently, the output productions produced by children were also analyzed, in order to make a comparison and understand if they correlate by quality and by quantity with input productions.

Several theoretical analyses on the ellipsis of various languages have been proposed in the literature. From the analysis of the results of previous studies on ellipsis, it is evident that young children are already quite capable of dealing with this linguistic phenomenon, and that children's direct speech contains many elliptical structures. However, the production of an elliptical structure does not imply understanding or full control of the structure itself, in fact, as the Corpus data show, it is evident that the use of ellipsis in children's direct speech can also lead to a misunderstanding between caregiver-child. While ellipsis is frequent in children's linguistic input and children produce some simple forms of ellipsis as early as the age of two, mastering all elliptical structures and being able to reconstruct previous linguistic information from speech may not be so easy for young children. In this regard, the purpose of data analysis based on the CHILDES Italian Tonelli Corpus was to test the role of the input children are exposed to in the acquisition of ellipsis.

In conclusion, data analysis has led to the claim that input productions containing ellipsis to which children are exposed during dialogue with caregivers partly influence output productions. Children, especially with increasing age, talk egocentrically on their own and from data analysis it was found that they produce ellipsis based on their own speech. For this reason, it can be argued that children from an early age begin to produce certain syntactic structures not only by imitation of the caregiver.

7

SUMMARY

The title of the dissertation is «Language acquisition process - Analysis on the production of *ellipsis* during caregiver-child dialogue».

The purpose of this dissertation is to delve into the language acquisition process, by testing the role of the input children are exposed to in the acquisition of syntax. The focus lies on the production of ellipsis in caregiver-child dialogues. This elaborate is divided into two parts: the first, consisting of three chapters, is dedicated to theoretical notions on the acquisition of language and ellipsis and the second one, composed by one chapter, is entirely reserved to the analysis of the data gathered from CHILDES Italian Tonelli Corpus.

The first chapter examines what language is and analyzes its distinctive features. Defining *language* is very complex, as it is an inflated term, to the point that there are many types of language, such as media language, language of young people, body language, language of animals. *Language*, in its most general sense, can be understood as a «symbolic system of communication». In particular, human language is a system of arbitrary vocal symbols by means of which the members of a society interact in terms of their total culture. Human language is unique among all forms of animal communication, it has its own characteristics, specific to the human species, which have been identified to distinguish it in particular from animal communication and in general from other semiotic systems.

Following this first introductory chapter, the second one focuses on the process of language acquisition, that is the process by which humans acquire the capacity to perceive and comprehend language, as well as to produce and use words and sentences to communicate. It is a very complex process, as it is characterized by very large individual variables that concern not only the times, but also the ways and strategies of learning. In this second chapter the main theories of language acquisition are explained, such as the behavioural and cognitive approaches and the emergent and innate theories.

The third chapter is dedicated to a review of the results of previous studies concerning the spontaneous ellipsis of children. Ellipsis is a rhetorical figure that consists in the omission, within a sentence, of one or more words that the grammatical construction would require and that it is possible to imply. From the analysis of the results of previous studies on ellipsis, it is evident that young children are already quite capable of dealing with this linguistic phenomenon and that children's direct speech contains many elliptical structures. In this regard, in the last

experimental chapter the different types of *remnant* of ellipsis used by caregivers and children in their conversations during game sessions will be analyzed and discussed. The aim is to verify whether children from an early age begin to produce certain syntactic structures due to imitation of the caregiver.

The collected data derive from the transcripts contained within a corpus, that is a collection of texts selected and organized to facilitate linguistic analyzes. In this case for data analyses was used the CHILDES Italian Tonelli Corpus, structured in three sub-corpora corresponding to three children Elisa, Gregorio and Marco, aged between one year and five months and two years and five months. Each sub-corpora contain the transcript of what the child and the caregiver, usually the mother but in Gregorio's case also the aunt, have produced during game sessions.

During the phase of data analysis what is technically called *remnant* was analyzed. The *remnant* is the lexicalized residual portion that remains after the ellipsis operation. Following the analysis of the types of ellipsis pronounced by caregivers, the question that has arisen is whether the quality and quantity of input from the caregiver correlate in quality and quantity with the child's output. To compare the output of the child to the input received, all remnant inputs pronounced by caregivers and all remnant outputs pronounced by children were entered into a single file in Excel. This work allowed to analyze each individual child in more detail, analyzing separately child input with child output and then examining input productions and output ones of each child by age.

This data analysis based on the Corpus has led to the claim that input productions containing ellipsis to which children are exposed during dialogue with caregivers partly influence output productions. Children, especially with increasing age, talk egocentrically on their own and from the analysis of the data it was found that they produce ellipsis based on their own speech. For this reason, it can be argued that children from an early age begin to produce certain syntactic structures not only by imitation of the caregiver.

In conclusion, acquiring language is something that happens naturally during the first years of life because human beings are predisposed to it, but it is a very complex process characterized by very large individual variables. For this reason, various studies and data analysis can offer their own results and conclusions useful to enrich the knowledge of the topic in question.

BIBLIOGRAFIA

- [1] Andorno Cecilia (2003). *Linguistica testuale. Un'introduzione*. Carocci, Roma.
- [2] Gillibrand Rachel, Lam Virginia, O'Donnell Victoria L., Tallandini Maria Anna (2019). *Psicologia dello sviluppo*. Pearson. Seconda edizione. Cap. 6.
- [3] Guasti Maria Teresa (2007). *L'acquisizione del linguaggio. Un'introduzione*. Raffaello Cortina Editore.
- [4] Lindenbergh Charlotte, Van Hout Angeliek, Hollebrandse Bart (2015). *Extending ellipsis research: The acquisition of sluicing in Dutch*. University of Groningen.
- [5] Ntelitheos Dimitris, Christodoulou Eleni (2005). *The acquisition of nominal ellipsis in greek*. UCLA Working Papers in Linguistics.
- [6] Sleeman Petra, Hulk Aafke (2013). *Acquisition of noun ellipsis in French and in Dutch – Consequences for linguistic theory*. ACLC, University of Amsterdam.
- [7] Versione digitale dell'Enciclopedia italiana Treccani di scienze, lettere e arti.
- [8] Website CHILDES Italian Tonelli Corpus con browsable transcripts
<https://childes.talkbank.org/access/Romance/Italian/Tonelli.html>
<https://sla.talkbank.org/TBB/childes/Romance/Italian/Tonelli>